ENGINEERING AND CHARACTERIZING RNA GENETIC REGULATORS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sarai Itagaki Meyer

January 2017

© 2017 Sarai Itagaki Meyer

ALL RIGHTS RESERVED

ENGINEERING AND CHARACTERIZING RNA GENETIC REGULATORS

Sarai Itagaki Meyer, Ph.D.

Cornell University 2017

The ability to predictably and accurately control gene expression is a fundamental requirement for synthetic biologists as they seek to harness the power of cellular machinery for real-world applications. To do so requires well-behaved genetic regulators that can be assembled to create the desired pattern of gene expression—be it the conditional translation of a reporter gene for molecular sensing or the phased production of pathway enzymes for the synthesis of a biopharmaceutical. I sought to engineer and characterize genetic regulators to address this need, focusing on regulators constructed of RNA because of its functional versatility, predictable folding, and ease of characterization. The research presented here covers the construction of RNA regulators designed to respond to external signals in the form of small molecules or light exposure, as well as the characterization of a naturally occurring RNA responsive to changes in temperature. In addition, I also report on the successful engineering and optimization of a series of synthetic small RNA transcriptional activators. Together, these projects expand the toolkit of genetic regulators available to synthetic biologists and inform future RNA engineering by improving our understanding of the RNA structure-function relationship.

BIOGRAPHICAL SKETCH

Sarai grew up in Worthington, Ohio and graduated from Thomas Worthington High School in 2007. She attended Yale University, where she conducted research in the Anastas Lab at the Center for Green Chemistry and Engineering and in the Wilson Lab in Chemical Engineering. She earned her B.S. in Engineering Sciences (Chemical) *cum laude* in May of 2011, then entered the PhD program in Cornell's School of Chemical and Biomolecular Engineering the following fall, eventually joining the lab of Julius Lucks in 2012.

To all those who've taught me along the way

ACKNOWLEDGEMENTS

Julius, thank you so much for your caring and thoughtful mentorship over the years and for your constant enthusiasm and refreshing candor. I'm incredibly grateful to have been a part of the Lucks Lab and for all your support and guidance. I can't thank you enough.

Matt and John, thank you for serving on my committee and for providing valuable advice and laboratory experience during my rotations.

To the Lucks Lab—past and present—thank you for making the lab a wonderful place and for all the advice, help, zaniness, comraderie, commiseration, and friendship you've provided over the my time here. Particular thanks go to Sitara and Rebecca for all their help in the research that went into this thesis and to Melissa for explaining everything back when I didn't know how to do anything. Alex, thank you for being my constant confidante, bubble tea buddy, and last-minute collaborator. Kyle, James, and Eric, thank you for being my experimental sounding boards on countless occasions. Katherine and Matt—I'm glad I got to know you over the last few months. I wish you all good luck on your new adventures!

To my friends, thank you for all the good times we've had together, for the work-outs, wine tastings, book clubs, and game nights. Thank you letting me rope you into volunteering for my latest event, for eating everything I cook (no matter how experimental), and for joining me in my often incompetent attempts at intramural athletics. In particular, Jonathan, thank you for your unique combination of cynicism and good cheer, for the wide-ranging

discussions, and for tolerating my frequent interruptions and silly jokes.

To my parents, thank you for encouraging my love of science and cheering me on through the rough patches. Poppy, I'm proud to be following in your footsteps.

Leah, thank you for always listening (even when I'm kvetching).

Michael, thank you for being my constant supporter in everything I do—I can't imagine doing it without you.

This work was supported by the National Science Foundation Graduate Research Fellowship Program (Grant Number 1144153).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The promise of engineering gene expression

Synthetic biology holds great promise for generating sophisticated solutions to address complex problems in human health, sustainability and biochemical production, but at its most fundamental level it depends largely on the successful manipulation of gene expression. For instance, microbes used to produce fuels and pharmaceuticals on an industrial scale require genetic engineering to express the appropriate pathway enzymes to generate the desired chemical product. Similarly, certain medical diagnostics depend upon the conditional gene expression of a reporter gene with a colored product in response to the presence of a pathogen RNA. To reliably control gene expression in this fashion, synthetic biologist rely on a variety of biological building blocks—some harvested from nature, others designed *de novo*—in order to assemble genetic networks that behave in a predictable fashion. However, the creation of increasingly complex circuitry requires a large toolbox of well-behaved biological parts with which to build these biological devices. My work seeks to tackle this growing need by generating RNA regulators for use in synthetic biology applications. While others within the synthetic biology community have focused on engineering protein regulators, we chose to work with RNA for its versatility of function, ease of design, and amenability to structural characterization (Chappell et al., 2013).

1.2 Why RNA?

Once viewed as simply an intermediary between DNA and protein, RNA has come to be recognized for its diverse roles in myriad cellular processes, particularly within the context of regulating gene expression (Chappell et al., 2013). Not only is RNA an incredibly versatile molecule, its simple four-letter alphabet and defined base-pairing rules make it a manageable target for rational design—allowing scientists to design RNAs that adopt desired conformations (Zadeh et al., 2011). Moreover, the pairing of RNA-structural characterization with next-generation sequencing (Loughrey et al., 2014; Lucks et al., 2011a) has enabled increasingly facile and accurate determination of RNA structure when combined with computational structure prediction. For these reasons we turned to RNA in our efforts to engineer gene expression for synthetic biology. The research reported here reflects my efforts to engineer various RNA regulators and to characterize a naturally occurring temperature-sensitive RNA regulator in order to better design future synthetic regulators. In order to give proper background for these results, I will first overview naturally occurring modes of genetic regulation, experimental methods of RNA structural determination, and recent advances in RNA synthetic biology before delving into the details of my research.

1.3 RNA genetic regulation in bacteria

Naturally occurring RNAs exhibit a vast array of functions that have only recently begun to be uncovered. Bacterial regulatory RNAs have been found to govern nearly every aspect of gene expression, from transcription to mRNA degradation to translation

(Chappell et al., 2013) (Figure 1.1), through mechanisms that will be briefly reviewed here, giving special attention to the regulators whose functions are relevant for later chapters.

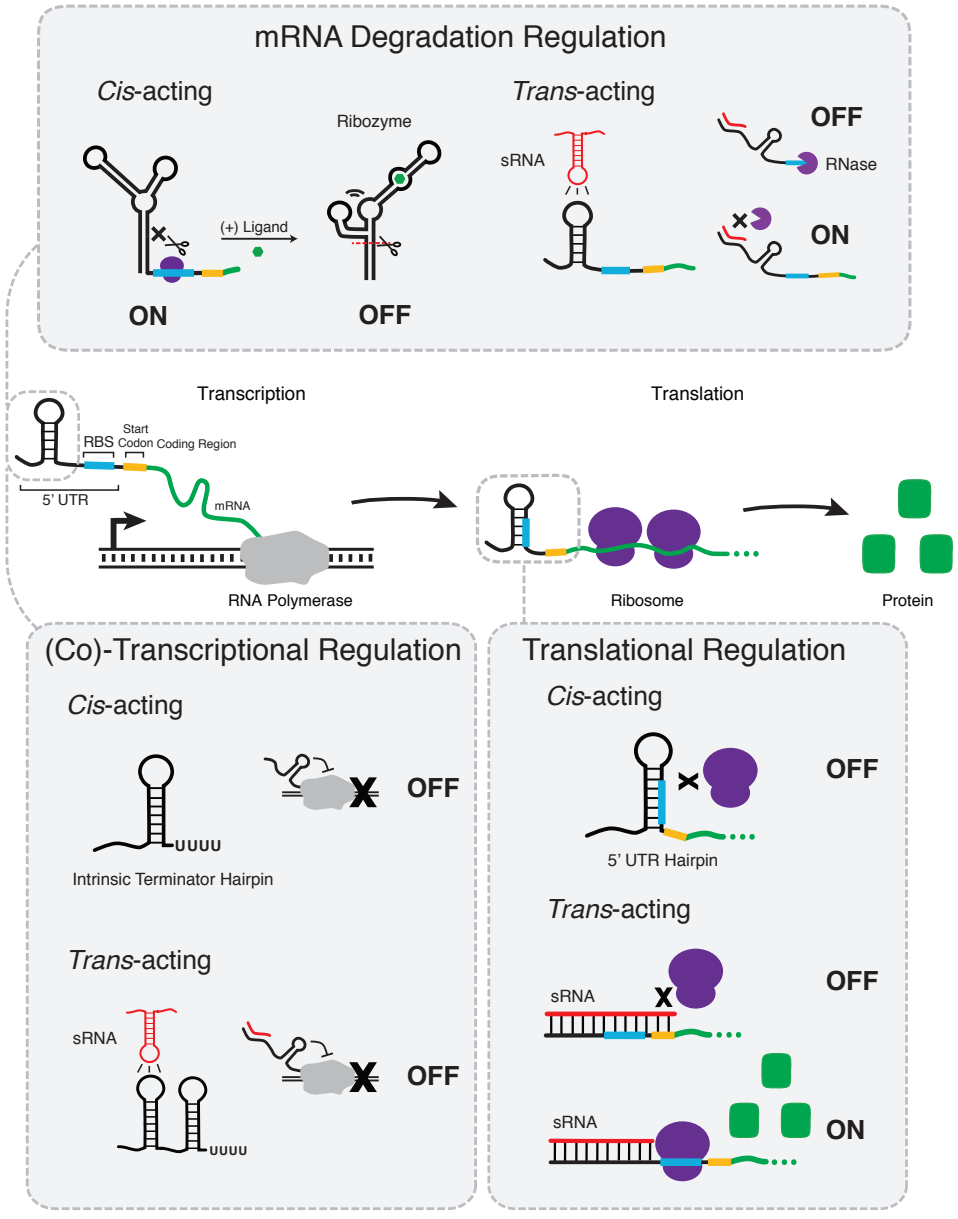


Figure 1.1 RNA regulation of gene expression. A schematic showing how naturally-occurring bacterial RNAs govern gene expression by regulating transcription, mRNA degradation, and translation. Figure adapted from works by Chappell et al. (2013, 2015).

1.3.1 Transcriptional control of gene expression

Since RNA folding occurs co-transcriptionally, RNA transcriptional regulation can act in *cis*, with the nascent RNA regulating its own elongation, or in *trans*, via the interaction of a separate sRNA with RNA being transcribed. RNA transcriptional regulation depends on the potential formation of intrinsic terminator hairpins within the 5' untranslated region (5' UTR) of the regulated gene. These hairpin sequences are followed by poly-uridine stretches that encourage polymerase pausing, allowing the hairpin to fold, thus causing the polymerase to abort transcription (Farnham and Platt, 1981; Larson et al., 2008). In some transcriptional riboswitches, like the *Bacillus subtilis* thiamine pyrophosphate (TPP) of flavin mononucleotide (FMN) riboswitches, the formation of this intrinsic terminator depends on the binding of a small ligand to an RNA aptamer domain (Mironov et al., 2002), while other riboswitches, like the glycine riboswitch, depend on ligand binding to disrupt terminator hairpin formation (Winkler et al., 2002).

Another regulatory mechanism of particular interest is transcriptional attenuation, a form of regulation that depends on the repressive action of a small RNA (sRNA) that targets an attenuator sequence in the 5' UTR of the regulated mRNA. In the presence of the sRNA, the attenuator forms an intrinsic terminator hairpin, prematurely aborting transcription, but in the absence of the sRNA, the attenuator folds into an interfered form, allowing transcription of the downstream gene (Brantl, 2007). Figure 1.2 illustrates this attenuation mechanism for the pT181 plasmid copy control element from *Staphylococcus aureus* (Brantl and Wagner, 2002).

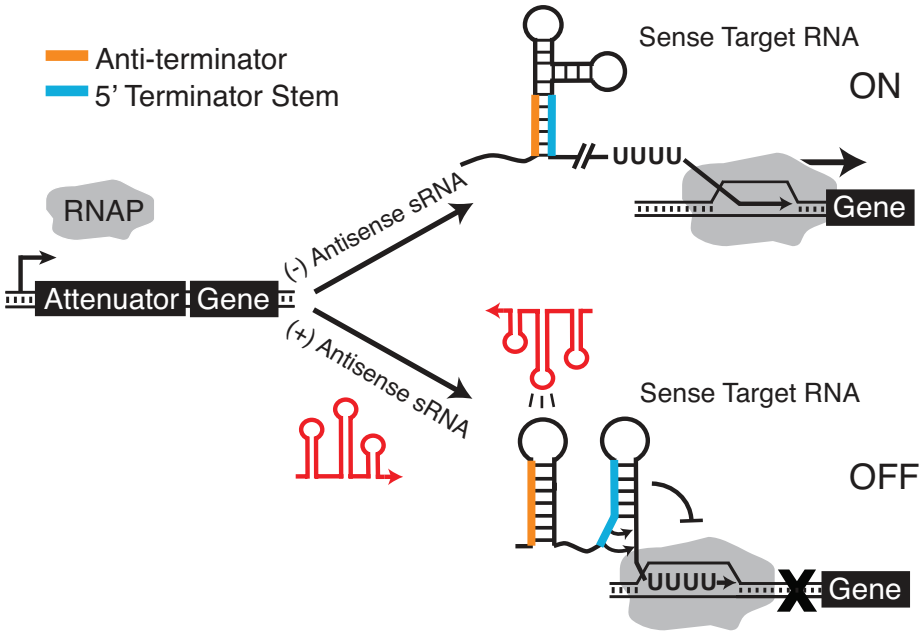


Figure 1.2 Mechanism of transcriptional repression of the pT181 attenuator. In the absence of the antisense sRNA, the anti-terminator region pairs with the 5' half of the terminator stem, allowing transcription (ON), but in the presence of antisense sRNA, the target forms the attenuator structure, causing intrinsic termination of transcription (OFF).

1.3.2 mRNA degradation control of gene expression

RNAs can also govern gene expression by regulating mRNA stability in a number of ways, either by directly causing RNA cleavage (in the case of ribozymes) or by triggering structural changes that allow or inhibit cleavage by RNase E, which initiates the main mode of mRNA degradation in bacteria (Carpousis, 2007). For ribozymes, direct RNA cleavage in response to a target ligand, as occurs in the *glmS* ribozyme upon glucosamine-6-phosphate binding, is followed by rapid breakdown of the cleaved fragments (Collins et al., 2007). In certain riboswitches like the *E. coli lysC* riboswitch, ligand-dependent structural rearrangements expose an RNase E cut site, allowing for mRNA degradation (Caron et al.,

2012). Other degradation control mechanisms rely on a trans-acting sRNAs to seed refolding events that prevent RNase cleavage (Opdyke et al., 2004), thereby stabilizing the mRNA, or encourage RNase E degradation, resulting in irreversible repression (Pfeiffer et al., 2009).

1.3.3 Translational control of gene expression

Lastly, RNAs can regulate gene expression by controlling the process of translation by influencing ribosome binding site (RBS) accessibility through a wide variety of mechanisms. Many sRNAs, function by directly binding to and blocking either the RBS or the start codon (Desnoyers et al., 2013), although others, like the *DsrA* sRNA in *E. coli*, activate translation by causing refolding events that expose the RBS (Majdalani et al., 1998) *Cis*-acting modes of RNA translational regulation include both activating and repressive riboswitches: for instance, the *S*-adenosylhomocysteine (SAH) riboswitch refolds to expose the RBS upon ligand binding (Wang et al., 2008), while the thiamine riboswitch occludes the RBS in response to binding of thiamine or thiamine pyrophosphate (Winkler et al., 2002). Another mechanism of interest relies on temperature-sensitive RNA structures that prevent RBS access at low temperatures while allowing translation at higher temperatures (Kortmann and Narberhaus, 2012)—these "RNA thermometers" will be discussed in Chapter 5 as one of the regulators I have chosen to study in greater detail.

1.4 RNA structural probing and prediction

Our ability to elucidate the mechanisms behind this vast array of RNA regulators is greatly aided by the existence of a broad suite of chemical and enzymatic probes to

determine RNA structure, combined with ever-improving methods for computational RNA structure prediction. Together, these two technologies enable increasingly accurate and high-thoughput determination of the structural features that underlie RNA's broad range of capabilities.

1.4.1 Experimental determination of RNA structures

There exist a wide range of methods for experimentally determining RNA structure, offering varying degrees of structural detail all the way down to the angstrom-level structural accuracy offered by X-ray crystallography. However, the highest resolution methods for determining RNA structure, namely X-ray crystallography and NMR (nuclear magnetic resonance) (Rinnenthal et al., 2010), require a high degree of specialization and are particularly low through-put, severely limiting the number of structures that can be determined via these strategies.

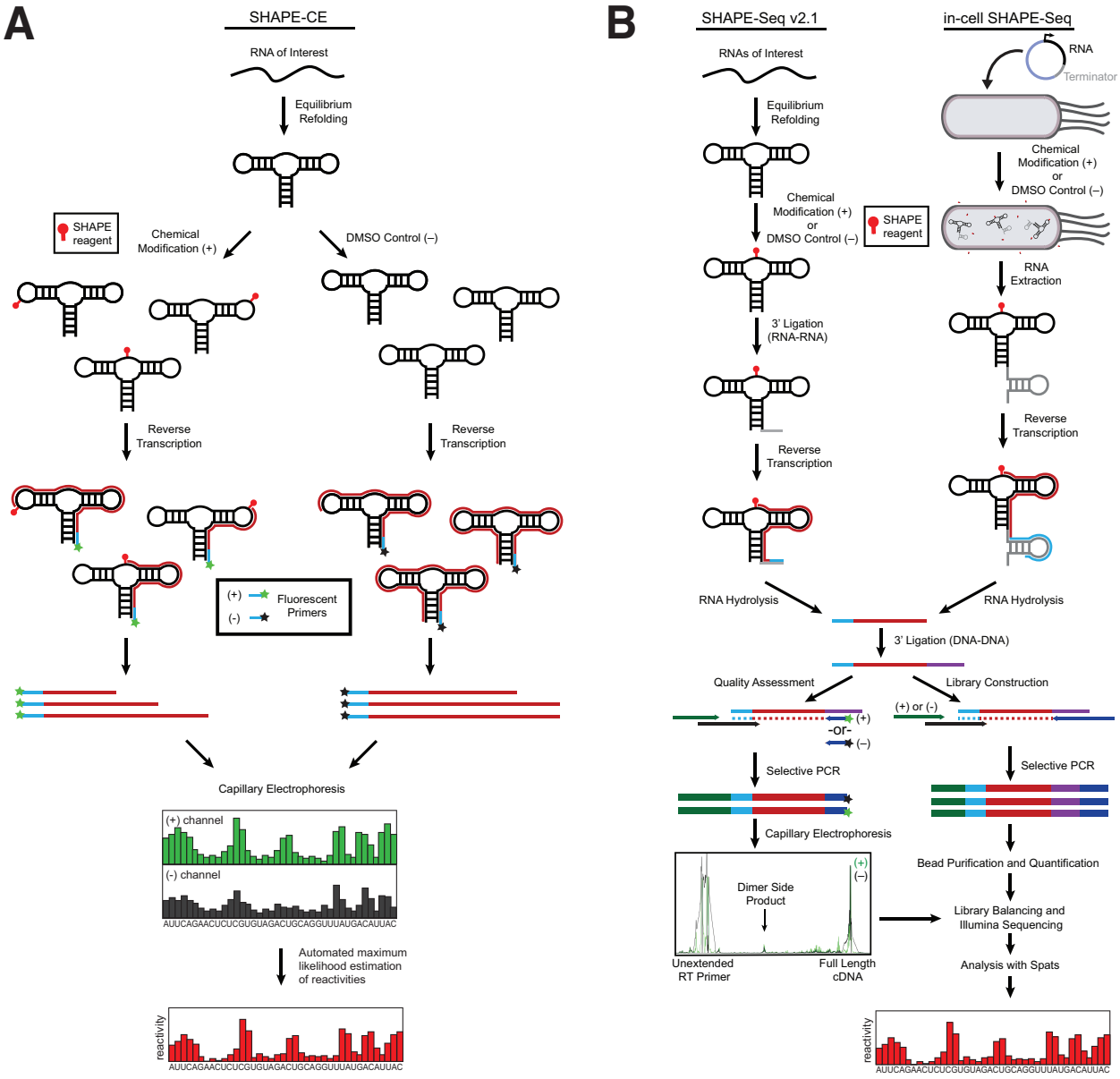
Other commonly used methods for determining RNA structure rely on either chemical or enzymatic probes to cleave or modify RNA in a structurally-dependent fashion (Knapp, 1989; Weeks, 2010). The locations of modification or cleavage can then be determined by either reverse transcription (halting at the site of modification) and sequencing or visualized via gel electrophoresis (Krol and Carbon, 1989). Combined with a knowledge of probe specificity—for instance, that RNase VI cleaves double-stranded regions, while RNase A cleaves at single-stranded pyrimidines—the read-outs of cleavage location can be used to infer aspects of the underlying structure (Knapp, 1989). Of particular utility are the family of chemical probes known as SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) reagents, including the widely used 1M7 (1-

methyl-7-nitroisatoic anhydride), which modify the 2'-hydroxyl of single-stranded nucleotides, allowing for sequence-agnostic probing of RNA structures at nucleotide resolution (Weeks and Mauger, 2011) with read-out via capillary electrophoresis through SHAPE-CE (Figure 1.3A). Not only have SHAPE probes been deployed in living cells (Spitale et al., 2013; Watters et al., 2015a), the method has also been paired with next-generation Illumina sequencing in SHAPE-Seq (Loughrey et al., 2014; Lucks et al., 2011a; Watters et al., 2015b) to allow for high-throughput structure determination of multiple RNAs simultaneously (Figure 1.3B), including monitoring the complete folding pathways of RNAs during *in vitro* transcription (unpublished data). Whether performed *in vitro* or *in* vivo, the output of each SHAPE-Seq experiment is a series of reactivities indicating the susceptibility of individual nucleotides to modification by SHAPE reagent (and thus their likelihood of being single-stranded within the RNA structural ensemble). This information can then be coupled with computational RNA structural prediction in order to arrive at a proposed RNA structure or structures (Hajdin et al., 2013; Leonard et al., 2013).

1.4.2 Computational prediction of RNA structures

RNA's logical base-pairing rules and limited alphabet have made it an attractive target for computational attempts at structural prediction. A number of algorithms have found reasonable success at predicting RNA secondary structures by calculating the structural partition function using it to determine the minimum free energy (MFE) structure, ViennaRNA (Hofacker, 2003) and RNAStructure (Reuter and Mathews, 2010) being two widely-used examples. Although algorithms of this type neglect non-canonical base-pairing and generally ignore tertiary structure, they can be made markedly more acc-

Figure 1.3 RNA structural probing via SHAPE. **(A)** Schematic outlining the key steps in SHAPE-CE probing of RNA structures. The RNA of interest is folded *in vitro*, subjected to chemical modification via SHAPE-reagent or a solvent control of DMSO (dimethyl sulfoxide), and then reverse transcribed using fluorescently labeled primers. The DNA fragments of different lengths are separated by capillary electrophoresis, and the resulting electropherogram peaks are analyzed to give the underlying nucleotide reactivities. **(B)** Schematic outlining RNA structure probing using SHAPE-Seq. SHAPE-Seq follows the same general strategy as SHAPE-CE, only the RNAs can be probed either *in vivo* or *in vitro*, and the DNA fragments are prepared for Illumina sequencing rather than capillary electrophoresis, requiring additional ligation and PCR steps to add flanking adapter sequences required by the Illumina sequencing platform. *Figure adapted from Watters et al. (2015b).


A
SHAPE-CE
RNA of Interest
Equilibrium Refolding
SHAPE reagent
Chemical Modification (+)
DMSO Control (–)
Reverse Transcription
Reverse Transcription
(+)
(-)
Fluorescent Primers
Capillary Electrophoresis
(+) channel
(-) channel
AUUCAGAACUCUCGUGUAGACUGCAGGUUUAUGACAUUAC
Automated maximum likelihood estimation of reactivities
reactivity
AUUCAGAACUCUCGUGUAGACUGCAGGUUUAUGACAUUAC
B
SHAPE-Seq v2.1
in-cell SHAPE-Seq
RNAs of Interest
RNA
Terminator
Equilibrium Refolding
Chemical Modification (+) or DMSO Control (–)
SHAPE reagent
Chemical Modification (+) or DMSO Control (–)
RNA Extraction
3' Ligation (RNA-RNA)
Reverse Transcription
Reverse Transcription
RNA Hydrolysis
RNA Hydrolysis
3' Ligation (DNA-DNA)
Quality Assessment
Library Construction
(+) or (-)
(+)
-or-
(–)
Selective PCR
Selective PCR
Capillary Electrophoresis
Bead Purification and Quantification
Dimer Side Product
(+)
(–)
Library Balancing and Illumina Sequencing
Unextended RT Primer
Full Length cDNA
Analysis with Spats
reactivity
AUUCAGAACUCUCGUGUAGACUGCAGGUUUAUGACAUUAC

-urate in by incorporating structural-probing data as a folding constraint (Hajdin et al., 2013; Leonard et al., 2013). In fact, using structural-probing data combined with computational prediction in this manner has proved useful for characterizing and designing engineered RNA regulators (Qi et al., 2012; Takahashi et al., 2016), making it a valuable tool for RNA synthetic biologists seeking to detailed structural information about their designed RNA constructs. Moreover, the need for such tools is increasing along with the recent growth of RNA synthetic biology—overviewed in the following section.

1.5 RNA synthetic biology—an overview

RNA synthetic biology has made great advances in recent years, with research in the last decade reporting the development of a variety of new RNA regulators as well as increasingly complex applications for synthetic RNAs, ranging from logic circuits and metabolic engineering to biosensors and molecular diagnostics (Chappell et al., 2015a). Many synthetic RNA regulators build off of pre-existing natural systems or use them as inspiration for their design: for instance, the naturally-occurring pT181 transcription regulator discussed previously (section 1.4.1) has proved a particularly fruitful launching point for RNA engineering—as part of a synthetic theophylline-responsive ncRNA (non-coding RNA) regulator (Qi et al., 2012), as a scaffold for building a library of orthogonal chimeric transcriptional attenuators (Takahashi and Lucks, 2013), or as a re-engineered regulator to activate transcription (Chappell et al., 2015b), as will be discussed in greater detail in Chapter 3. Other naturally-occurring RNA regulators have inspired synthetic variants as well, from synthetic riboswitches created by fusing aptamer domains to pre-existing riboswitch expression platforms (Ceres et al., 2013a, 2013b) to synthetic sRNAs to

regulate mRNA stability and degradation (Man et al., 2011) and synthetic RNA thermosensors that expose RNase E sites for degradation in response to changes in temperature (Hoynes-O'Connor et al., 2015). Other synthetic RNA regulators include the riboregulator developed by Isaacs et al. (2004), which represses translation in *cis* through an RNA hairpin designed to block access to the RBS unless a *trans*-acting sRNA binds and prevents hairpin formation. This design strategy has been simplified with the creation of toehold switches (Green et al., 2014) (Figure 1.4), which allow for efficient translation activation and computational design via the NUPACK design algorithm (Zadeh et al., 2010).

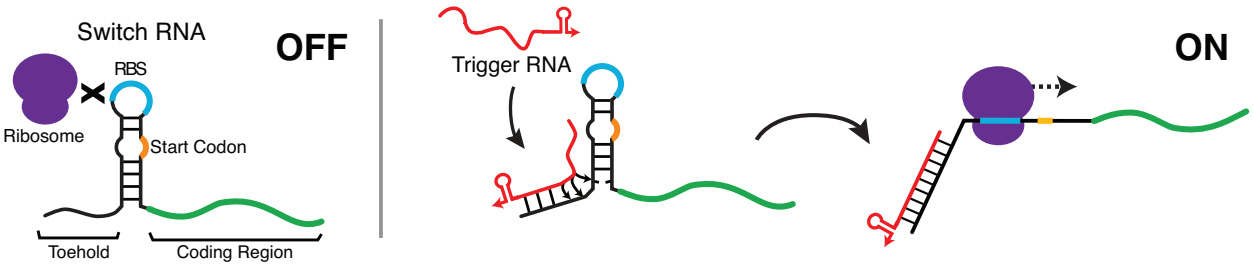


Figure 1.4 Mechanism of the RNA toehold switch. In the OFF state, the RBS is inaccessible in a hairpin loop, but the binding of the designed trigger RNA to the toehold sequence unzips the adjacent hairpin, allowing for ribosome access and translation. *Figure adapted from Chappell et al. (2015a).

Many of these synthetic regulators have been used to build RNA circuits, including RNA logic gates (networks which perform basic digital computations) using toehold switches (Green et al., 2014), an AND gate constructed using small transcriptional activating RNAs (Chappell et al., 2015b), and a NOR gate constructed from ligand-sensing

transcriptional regulators (Qi et al., 2012). NOR gates have also been constructed using CRISPR interference (CRISPRi), in which a catalytically dead version of Cas9 (dCas9) is paired with a small guide RNA (sgRNA) that targets a particular mRNA to repress transcription (Nielsen and Voigt, 2014). Extending their application beyond basic logic gates, catalytically dead CRISPRs have also been fused with light-responsive proteins to create optogenetic regulators targeted using engineered sgRNAs (Nihongaki et al., 2015; Polstein and Gersbach, 2015). In other applications of note for RNA synthetic biology, ribosome binding site engineering based on the RBS Library Calculator has been used for metabolic engineering, optimizing flux through a biosynthetic pathway (Farasat et al., 2014), and toehold switches have been applied to create paper-based molecular diagnostics capable of detecting mRNAs and distinguishing between sequences from the Zaire and Sudan strains of the Ebola virus (Pardee et al., 2014). Rather than using cells, these impressive paper-based biosensors rely on freeze-dried cell-free transcription-translation systems to generate a colorimetric output. As outlined below, these cell-free systems have broad utility for synthetic biology.

1.6 The TX-TL cell extract system and its utility

Cell-free transcription-translation systems have long been used as a way to express proteins outside of the cellular environment (Hoagland et al., 1958; Wood and Berg, 1962). Recent refinements in the preparation of these cell extracts have improved their performance (Kigawa et al., 2004; Sun et al., 2013), and the development of transcription-translation systems built from carefully assembled purified components (Shimizu et al., 2001) has allowed for the introduction of non-canonical amino acids and the careful

dissection of transcriptional mechanisms (Mishler and Gallivan, 2014). In transcription-translation (TX-TL) systems, any nucleic acid sequence or molecule of interest can be added, without regard to cell uptake or membrane permeability, and these components can be added at any point during the reaction, allowing for the easy monitoring of the system's response. Moreover, cell-free TX-TL systems shorten the design-build-test cycle for developing genetic circuit elements by eliminating the need to transform cells with the new genetic constructs of interest, reducing the time required for testing from 3 days to 2-3 hours and making TX-TL an especially attractive technology for synthetic biologists.

In broad strokes, the production of TX-TL cell extract requires the preparation of a crude *E. coli* cell lysate that goes through a series of treatment steps—centrifugation, digestion, and dialysis—in order to remove unwanted components from the lysate. For experimental use, the cell lysate is combined with a buffer solution of amino acids, nucleotides, co-factors, and batch-specific optimized levels of magnesium-glutamate, potassium-glutamate, and dithiothreitol (Sun et al., 2013). DNA templates (plasmid or linear), RNAs, and small molecules can be added directly to the TX-TL reactions, which are generally carried out in 384-well plates and monitored via plate reader.

Early work establishing the viability of genetic circuit assembly in cell-free systems demonstrated the use of both EGFP and firefly luciferase as reporter genes and validated the function of a multi-level cascades and genetic circuits containing both positive and negative regulatory elements (Noireaux et al., 2003). Since then, TX-TL cell-free systems have been used to prototype artificial cells (Noireaux and Libchaber, 2004), to pattern gene expression to emulate *Drosophila* developmental gene networks (Isalan et al., 2005), and to completely synthesize the T7 bacteriophage (Shin et al., 2012). However, from the

standpoint of synthetic biology, a key strength of the TX-TL cell-free system remains its efficiency for quickly testing genetic constructs without the need to undergo bacterial transformation, allowing TX-TL to function as a cell-free breadboard for prototyping genetic regulators and circuits (Siegal-Gaskins et al., 2014; Sun et al., 2014). In particular, recent work by Takahashi et al. (2015) used TX-TL to measure the response time of an RNA transcriptional cascade and test an RNA single input module shown to function *in vivo*, demonstrating the clear utility of the cell-free TX-TL system for testing RNA genetic regulators.

1.7 Engineering and understanding RNA regulators

Motivated by the need for novel genetic regulators to expand the synthetic biology toolkit and the clear benefits of working with RNA, I set out to build and characterize a series of RNA genetic regulators.

1.7.1 Chapter 2: Toward a light-activated RNA regulator

One of the key goals of synthetic biology is to engineer biological systems capable of sensing and responding to their environment in predictable fashion. In particular, light offers interesting opportunities for facile temporal and spatial control of external signaling, making it an attractive potential input for genetic circuits. Most previous optogenetic synthetic biology applications to date have relied upon light-sensing proteins (Gardner and Deiters, 2012), but I sought to build off previous research on light-sensitive RNA aptamers (Hayashi et al., 2009; Lee et al., 2007; Young and Deiters, 2008) in order to create a novel, light-sensitive RNA transcriptional regulator. My attempts relied on RNA aptamers that

selectively bound one form of a small molecule or peptide that underwent light-catalyzed structural isomerization, with the goal that this binding signal would then be transduced to a fused pre-existing regulator, either the pT181 attenuator system (Lucks et al., 2011b) or a riboswitch expression platform (Ceres et al., 2013a). This strategy for generating light-sensing RNA regulators ultimately proved unsuccessful, but the attendant study of aptamer-ligand binding reaffirmed the utility of SHAPE probing for identifying regions of ligand binding in RNA aptamers.

1.7.2 Chapter 3: Improving fold activation of small RNAs with rational engineering strategies

The recent advent of small transcription activating RNAs (STARs) (Chappell et al., 2015b) (mentioned in section 1.2) filled an important gap in RNA regulation, giving synthetic biologists the ability to activate as well as repress gene expression on the transcriptional level using small RNAs. These newly developed regulators allowed for novel RNA circuit architectures; however, their utility was limited by the small number of highly functional STARs available for construction of complex circuits. Our research addressed this limitation by generating additional highly functional STARs through a series of rational RNA engineering strategies (Meyer et al., 2016). We found that the addition of 5' stability hairpins and 3' sRNA scaffolds to STARs, combined with increasing the ratio of STAR antisense to target sense, resulted in large increases in fold-activation for a number of low-performing STARs. Moreover, the newly optimized STARs were largely orthogonal to each other and to pre-existing regulators, making them broadly useful for future RNA circuitry.

1.7.3 Chapter 4: Engineering ligand-responsive RNA activators

Building off of the invention of STARs, I sought to generate a new mechanism of RNA regulation: ligand-activated transcriptional activation in *trans* via aptamer-STAR fusions or "aptaSTARs." Using a combination of rational design and random-screening approaches, I was able to generate a few ligand-sensitive aptaSTARs for the small-molecule theophylline. However, the aptaSTARs lacked the dynamic range of their original parent STARs, making them less useful as potential circuit elements, and hinting at remaining gaps in our ability to design switchable RNA structures.

1.7.4 Chapter 5: Characterizing the structure-function relationship of a naturally occurring RNA thermometer

An intriguing example of RNA-based gene regulation, the *agsA* RNA thermometer from *Salmonella enterica* allows the bacterium to sense the temperature change associated with entering the human gut and up-regulate the AgsA heat-shock protein in response. Using a combination of *in vivo* and *in vitro* methods, I characterized the structure and function of the *agsA* regulator, establishing that its mechanism is based on an equilibrium-based melting of the helix containing the ribosome binding site. Moreover, I demonstrated the ability of SHAPE-Seq to detect reactivity changes within the *agsA* thermometer caused by ribosome binding and the resultant structural changes.

1.7.5 Chapter 6: Conclusions and perspectives

As demonstrated in the presented work, RNA regulators provide broad potential for genetic regulation and synthetic biology. I have succeeded in engineering new and improved RNA regulators and in uncovering the mechanism behind those found in nature. Through improving our understanding between the relationship between RNA structure and function, this work will better our ability to design new regulators for diverse synthetic biology applications. It is my hope that future research will further enhance our understanding of RNA engineering and enable the design of increasingly sophisticated RNA circuitry able to properly address complex problems.

1.8 Introduction bibliography

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr. Opin. Microbiol. *10*, 102–109.

Brantl, S., and Wagner, E.G.H. (2002). An Antisense RNA-Mediated Transcriptional Attenuation Mechanism Functions in Escherichia coli. J. Bacteriol. *184*.

Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. Proc. Natl. Acad. Sci. U. S. A. *109*, E3444–E3453.

Carpousis, A.J. (2007). The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. Annu. Rev. Microbiol. *61*, 71–87.

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013a). Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth. Biol. *2*, 463–472.

Ceres, P., Trausch, J.J., and Batey, R.T. (2013b). Engineering modular “ON” RNA switches using biological components. Nucleic Acids Res. *41*, 10449–10461.

Chappell, J., Takahashi, M.K., Meyer, S., Loughrey, D., Watters, K.E., and Lucks, J. (2013). The centrality of RNA for engineering gene expression. Biotechnol. J. *8*, 1379–1395.

Chappell, J., Watters, K.E., Takahashi, M.K., and Lucks, J.B. (2015a). A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. Curr. Opin. Chem. Biol. *28*, 47–56.

Chappell, J., Takahashi, M.K., and Lucks, J.B. (2015b). Creating small transcription activating RNAs. Nat. Chem. Biol. *11*, 1–9.

Collins, J.A., Irnov, I., Baker, S., and Winkler, W.C. (2007). Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev. *21*, 3356–3368.

Desnoyers, G., Bouchard, M.-P., and Massé, E. (2013). New insights into small RNA-dependent translational regulation in prokaryotes. Trends Genet. *29*, 92–98.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H.M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol. Syst. Biol. *10*, 731.

Farnham, P.J., and Platt, T. (1981). Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. Nucleic Acids Res. *9*, 563–577.

Gardner, L., and Deiters, A. (2012). Light-controlled synthetic gene circuits. Curr. Opin. Chem. Biol. *16*, 292–299.

Green, A.A., Silver, P.A., Collins, J.J., and Yin, P. (2014). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. Cell *159*, 1–15.

Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H., and Weeks, K.M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc. Natl. Acad. Sci. *110*, 5498–5503.

Hayashi, G., Hagihara, M., Nakatani, K., Dohno, C., and Nakatani, K. (2009). RNA aptamers that reversibly bind photoresponsive azobenzene-containing peptides. Chemistry (Easton). *15*, 424–432.

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., and Zamecnik, P. (1958). A soluble ribonucleic acid intermediate in protein synthesis. J. Biol. Chem. *231*, 241–257.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Hoynes-O'Connor, A., Hinman, K., Kirchner, L., and Moon, T.S. (2015). De novo design of heat-repressible RNA thermosensors in E. coli. Nucleic Acids Res. 1–14.

Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. *22*, 841–847.

Isalan, M., Lemerle, C., and Serrano, L. (2005). Engineering gene networks to emulate Drosophila embryonic pattern formation. PLoS Biol. *3*, 0488–0496.

Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., and Yokoyama, S. (2004). Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. J Struct Funct Genomics *5*, 63–68.

Knapp, G. (1989). Enzymatic Approaches to Probing of RNA Secondary and Tertiary Structure. Methods Enzymol. *180*, 192–212.

Kortmann, J., and Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers

and switches. Nat. Rev. Microbiol. *10*, 255–265.

Krol, A., and Carbon, P. (1989). A Guide for Probing Native Small Nuclear RNA and Ribonuclear Protein Structures. Methods Enzymol. *180*, 212–226.

Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. Cell *132*, 971–982.

Lee, H.-W., Robinson, S.G., Bandyopadhyay, S., Mitchell, R.H., and Sen, D. (2007). Reversible photo-regulation of a hammerhead ribozyme using a diffusible effector. J. Mol. Biol. *371*, 1163–1173.

Leonard, C.W., Hajdin, C.E., Karabiber, F., Mathews, D.H., Favorov, O. V, Dokholyan, N. V, and Weeks, K.M. (2013). Principles for Understanding the Accuracy of SHAPE-Directed RNA Structure Modeling. Biochemistry *52*, 588–595.

Loughrey, D., Watters, K.E., Settle, A.H., and Lucks, J.B. (2014). SHAPE-Seq 2.0 : systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. *42*, 1–10.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011a). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D., and Arkin, A.P. (2011b). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc. Natl. Acad. Sci. *108*, 8617–8622.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism , independent of its

action as an antisilencer of transcription. Proc. Natl. Acad. Sci. *95*, 12462–12467.

Man, S., Cheng, R., Miao, C., Gong, Q., Gu, Y., Lu, X., Han, F., and Yu, W. (2011). Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. Nucleic Acids Res. *39*, e50.

Meyer, S., Chappell, J., Sankar, S., Chew, R., and Lucks, J.B. (2016). Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. Biotechnol. Bioeng. *113*, 216–225.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002). Sensing Small Molecules by Nascent RNA : A Mechanism to Control Transcription in Bacteria. Cell *111*, 747–756.

Mishler, D.M., and Gallivan, J.P. (2014). A family of synthetic riboswitches adopts a kinetic trapping mechanism. Nucleic Acids Res. *42*, 6753–6761.

Nielsen, A.A.K., and Voigt, C.A. (2014). Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. Mol. Syst. Biol. *10*, 763.

Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., and Sato, M. (2015). CRISPR-Cas9-based photoactivatable transcription system. Chem. Biol. *22*, 169–174.

Noireaux, V., and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. Proc. Natl. Acad. Sci. U. S. A. *101*, 17669–17674.

Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003). Principles of cell-free genetic circuit assembly. Proc. Natl. Acad. Sci. U. S. A. *100*, 12672–12677.

Opdyke, J.A., Kang, J., and Storz, G. (2004). GadY, a Small-RNA Regulator of Acid Response Genes in Escherichia coli GadY. J. Bacteriol. *186*, 6698–6705.

Pardee, K., Green, A.A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., and Collins, J.J.

(2014). Paper-Based Synthetic Gene Networks. Cell *159*, 940–954.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C.D., and Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. Mol. Biol. *16*, 840–846.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat. Chem. Biol. *11*, 198–200.

Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K., and Arkin, A.P. (2012). Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res. *40*, 5775–5786.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rinnenthal, J., Klinkert, B., Narberhaus, F., and Schwalbe, H. (2010). Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. Nucleic Acids Res. *38*, 3834–3847.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755.

Shin, J., Jardine, P., and Noireaux, V. (2012). Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-Free reaction. ACS Synth. Biol. *1*, 408–413.

Siegal-Gaskins, D., Tuza, Z.A., Kim, J., Noireaux, V., and Murray, R.M. (2014). Gene circuit performance characterization and resource usage in a cell- free “breadboard.” ACS Synth. Biol. *3*, 416–425.

Spitale, R.C., Crisalli, P., Flynn, R. a, Torre, E. a, Kool, E.T., and Chang, H.Y. (2013). RNA

SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J. Vis. Exp. *79*, e50762.

Sun, Z.Z., Yeung, E., Hayes, C.A., Noireaux, V., and Murray, R.M. (2014). Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. ACS Synth. Biol. *3*, 387–397.

Takahashi, M.K., and Lucks, J.B. (2013). A modular strategy for engineering orthogonal chimeric RNA transcription regulators. Nucleic Acids Res. *41*, 7577–7588.

Takahashi, M.K., Chappell, J., Hayes, C.A., Sun, Z.Z., Kim, J., Singhal, V., Spring, K.J., Al-Khabouri, S., Fall, C.P., Noireaux, V., et al. (2015). Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. ACS Synth. Biol. *4*, 503–515.

Takahashi, M.K., Watters, K.E., Gasper, P.M., Abbott, T.R., Carlson, P.D., Chen, A.A., and Lucks, J.B. (2016). Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. RNA *22*, 920–933.

Wang, J.X., Lee, E.R., Morales, D.R., Lim, J., and Breaker, R.R. (2008). Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol. Cell *29*, 691–702.

Watters, K.E., Abbott, T.R., and Lucks, J.B. Simultaneous Characterization of Cellular RNA Structure and Function with in-cell SHAPE-Seq.

Watters, K.E., Abbott, T.R., and Lucks, J.B. (2015a). Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. gkv879.

Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H., and Lucks, J.B. (2015b). Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). METHODS 1–37.

Weeks, K.M. (2010). Advances in RNA structure analysis by chemical probing. Curr. Opin. Struct. Biol. *20*, 295–304.

Weeks, K.M., and Mauger, D.M. (2011). Exploring RNA Structural Codes with SHAPE Chemistry. Acc. Chem. Res. *12*, 1280–1291.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.

Wood, W., and Berg, P. (1962). The effect of enzymatically synthesized ribonucleic acid on amino acid incorporation by a soluble protein-ribosome system from Escherichia coli. Proc. Natl. Acad. Sci. *4*, 94–104.

Young, D.D., and Deiters, A. (2008). Light-regulated RNA-small molecule interactions. Chembiochem *9*, 1225–1228.

Zadeh, J.N., Wolfe, B.R., and Pierce, N.A. (2010). Nucleic Acid Sequence Design via Efficient Ensemble Defect Optimization. J. Comput. Chem. *32*, 439–451.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: analysis and design of nucleic acid systems. J. Comput. Chem. *32*, 170–173.

CHAPTER 2

TOWARD A LIGHT-ACTIVATED RNA REGULATOR

2.1 Abstract

Making devices that sense and respond to signals in their environments is a key goal of synthetic biology. In particular, light represents an attractive and easy way to interface with biological machinery—enabling exquisite spatial and temporal control and essentially instantaneous sensing, thus avoiding the delays for diffusion and uptake often associated with molecular signals. This work details a series of attempts to engineer a light-sensitive RNA switch in order to expand the synthetic biology toolbox and allow for light-based gene control at the transcriptional level. While it did not succeed in generating a light-activated switch, the reported research did help establish the utility of the SHAPE technique for identifying the location of ligand-binding in an RNA aptamer.

2.2 Introduction

Realizing the full potential of synthetic biology will involve creating biological devices capable of sensing and responding to environmental cues, and including light in the suite of detectable signals offers a number of obvious advantages. Not only do light signals allow for accurate spatial and temporal control, they can also be sensed virtually instantaneously and are incredibly cost-effective in comparison to expensive molecular inducers.

The dependence of nearly all earthly life on harnessing energy from sunlight has resulted in a vast array of naturally occurring protein-based mechanisms for sensing and responding to light. Scientists have co-opted many of these proteins for a wide variety of applications, ranging from neuronal control and monitoring (Song and Knöpfel, 2015) to

the creation of light-controlled synthetic gene circuits (Gardner and Deiters, 2012). In particular, the opsin protein family has found widespread use as a neuronal optogenetic tool (Zhang et al., 2011), allowing researchers to directly photostimulate mammalian neurons (Boyden et al., 2005), drive movement preference in mice through *in vivo* control of intracellular signaling (Airan et al., 2009), and detect individual action potentials in cultured rat brain cells (Kralj et al., 2012).

Within the field of synthetic biology, naturally occurring light-sensing proteins have been employed and engineered for a broad range of uses from basic genetic regulation to the creation of complex devices. Various light-activated gene regulation mechanisms have been developed to act at the transcriptional and translational levels. In one instance, Cao et al. (2013) engineered a platform for light-induced mRNA translation by fusing a eukaryotic translation initiation factor to the photo-sensitive *Arabodopsis* cryptochrome protein CRY2 and an RNA binding domain to its illumination-dependent binding partner, CIB1. Following a different strategy, Ryu and Gomelsky (2014) used an engineered light-activated bacteriophytochrome diguanylate cyclase to regulate production of a signaling molecule c-di-GMP, thereby controlling a c-di-GMP-dependent transcription factor and target gene.

Research by Tabor and co-workers took this idea of light-controlled gene expression one step further in order to generate networks capable of edge-detection and bacterial photography. By generating a chimeric light receptor (Cph8) composed of a photoreceptor fused to the EnvZ-OmpR histidine kinase and response-regulator, they created a system in which gene expression from the OmpC promoter was prevented in the presence of red light. Then, by placing the *LacZ* reporter gene under the OmpC promoter, they generated

bacterial "film" capable of recording a pattern of red light projected onto S-gal containing media (Levskaya et al., 2005). When this same chimeric light receptor was used to regulate the intercellular signal AHL (3-oxohexanoyl-homoserine lactone) and combined with clever circuit architecture for expression of *LacZ* upon AHL detection in the absence of light, the result was a bacterial edge-detection system, capable of printing in cleaved S-gal the outline of a projected photomask (Tabor et al., 2009). Other light-sensitive applications of note include the use of light-induced protein interactions to trigger membrane-localization of a desired target, allowing for light-directed lamellipodial extension in mammalian cells (Levskaya et al., 2005), and the use of light-induced transgene expression to regulate insulin levels in diabetic mice (Ye et al., 2011).

Because of the inherent difficulty of generating novel light-sensitive proteins (Levskaya et al., 2009), all of the work discussed so far relied on pre-existing light-responsive proteins and chromophores, generally fused to another domain or protein with the desired activity. However, a few groups have investigated ways to use photo-caging to introduce light-sensitivity to new proteins or interactions. Recent research has demonstrated the generation of newly photo-sensitive proteins through the incorporation of amino acids modified with photo-labile protecting groups. Photo-caged serine was used to create a light-sensitive variant of the transcription factor Pho4 by blocking phosphorylation of key residues whose phosphorylation state determines Pho4's nuclear export kinetics (Lemke et al., 2007). Similar work by Gautier et al. (2011) introduced a photo-caged lysine into a MAP kinase to sterically block ATP-binding in a light-dependent fashion. Photo-caging has also been applied to small-molecule targets like IPTG (Young and

Deiters, 2007) and doxycycline (Cambridge et al., 2006) as an alternate strategy for introducing light-sensitivity to a pre-existing system.

Although RNA-based regulators offer relative ease of engineering and selection, in comparison to the explosion of research on protein-based optogenetics, little research has focused on building light-sensitive RNA devices, largely because of the lack of naturally-occurring light-sensitive RNAs (Jäschke, 2012). However, a few key papers have explored the possibility of generating light-responsive RNAs by taking advantage of RNA's ability to bind a wide variety of targets with high specificity. Here I give a brief overview of the relevant work to date before launching into a discussion of our attempts to generate a light-activated RNA regulator by building off of previous research.

Just as photo-responsive proteins use bound chromophores to sense light, photo-responsive RNAs depend on small molecules that undergo defined structural changes in response to light. For instance, a dihydropyrene compound known as BDHP-COOH (10-carboxy-2,7-di-t-butyl-*trans*-12c,12d-dimethyl-12c,12d-dihy-drobenzo[e]pyrene) that isomerizes into its "open" form (BCPD-COOH) in response to visible light (Figure 2.1A) was used to generate a photo-responsive hammerhead ribozyme (Lee et al., 2007a). Using systematic evolution of ligands by exponential enrichment (SELEX, see section A.2 for background) (Ellington and Szostak, 1990; Tuerk and Gold, 1990), they isolated an aptamer with high specificity for the closed BDHP form of the target dihyropyrene. Next, they identified the minimal aptamer necessary for BDHP binding by assaying the binding affinity of a series of truncated aptamers. This minimal BDHP aptamer was then fused to the hammerhead ribozyme via a handful of potential communication modules, and the resulting suite of ribozymes were tested for light and ligand-dependent cleavage. A

construct dubbed the UGLOOP hammerhead ribozyme (Figure 2.1B) showed clear BDHP-dependent cleavage and irradiation-dependent activity, demonstrating the successful creation of a light-dependent ribozyme mediated by ligand interaction.

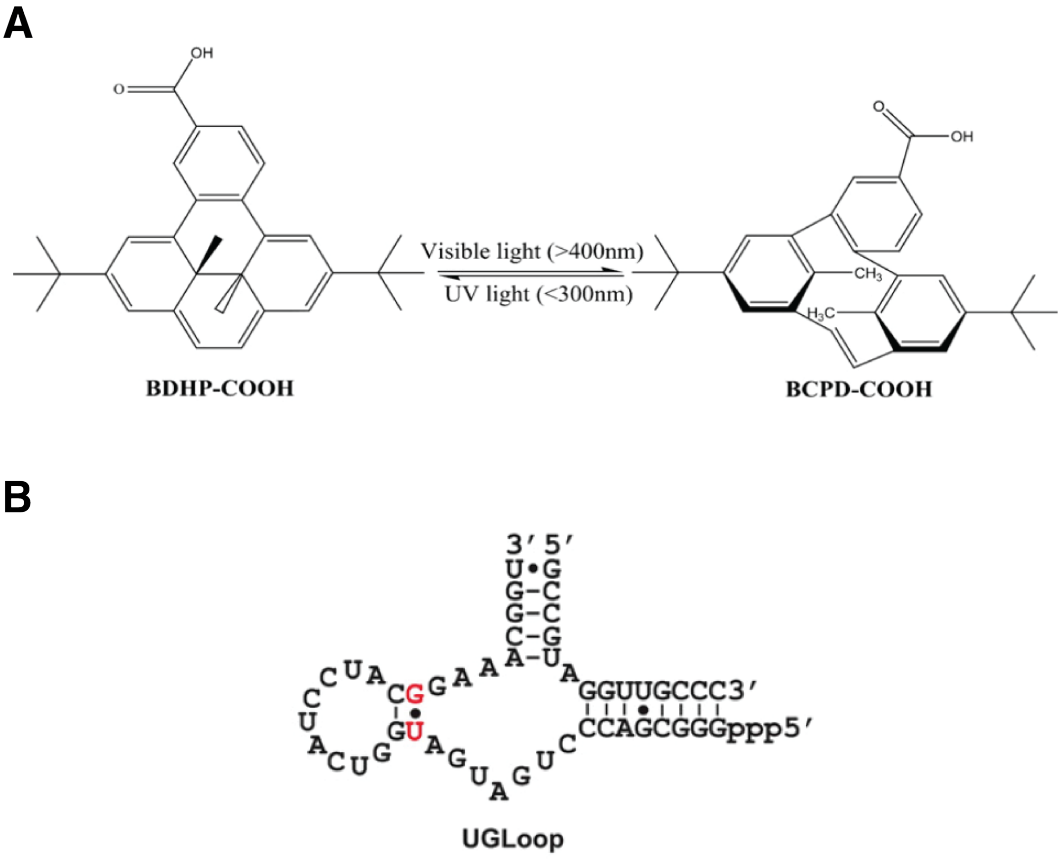


Figure 2.1 A light-responsive dihydropyrene-binding ribozyme.* **(A)** Isomerization between the closed BDHP-COOH form and open BCPD-COOH form of 10-carboxy-2,7-di-t-butyl-*trans*-12c,12d-dimethyl-12c,12d-dihy-drobenzo[e]pyrene. **(B)** Putative structure of the UGLOOP hammerhead ribozyme construct whose cleavage activity showed clear dependence on the dihydropyrene's BDHP form. *Figures reprinted from Lee et al. (2007).

The small molecule spiropyran also undergoes a photocatalyzed isomerization, transitioning between its normal colorless state and its purple merocyanine form in response to visible or UV light (Figure 2.2). Research by Asanuma et al. (2001) had taken advantage of this structural change to generate a spiropyran-tethered DNA strand with a light-modulated duplex melting temperature. More recent work by Young and Deiters (2008) used SELEX to isolate an aptamer specific to the spiropyran form via carefully alternated positive and negative selection. The aptamer pool was incubated with a resin containing bound spiropyran, and non-binding aptamers were washed away. The resin was then exposed to UV light to catalyze the formation of the merocyanine form, and the

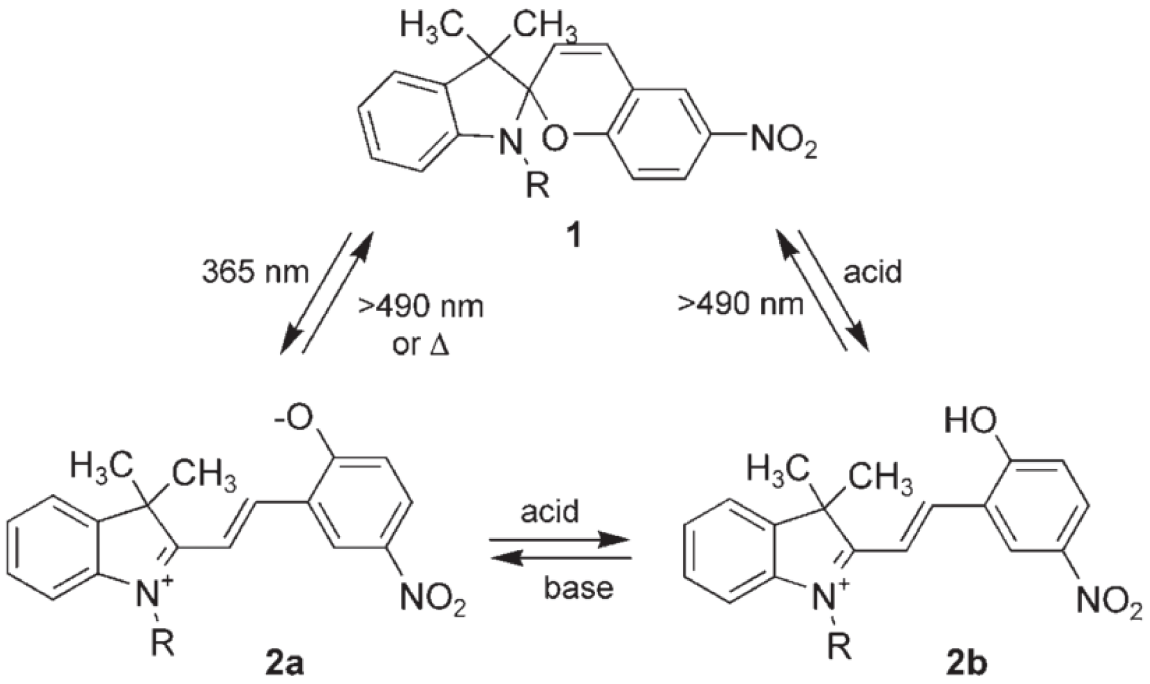


Figure 2.2 The structural transitions of the light-sensitive spiropyran molecule.* The colorless spiropyran form **(1)** transitions between its purple merocyanine form **(2a)** and yellow merocyanine form **(2b)** in response to exposure to UV light or changes in pH. *Reprinted from Young and Deiters (2008) with permission from John Wiley and Sons.

aptamers that did not bind were eluted and collected. After ten rounds of SELEX, an aptamer variant with good specificity for spiropyran was identified, and its reversible binding to immobilized spiropyran was confirmed using surface plasmon resonance (SPR).

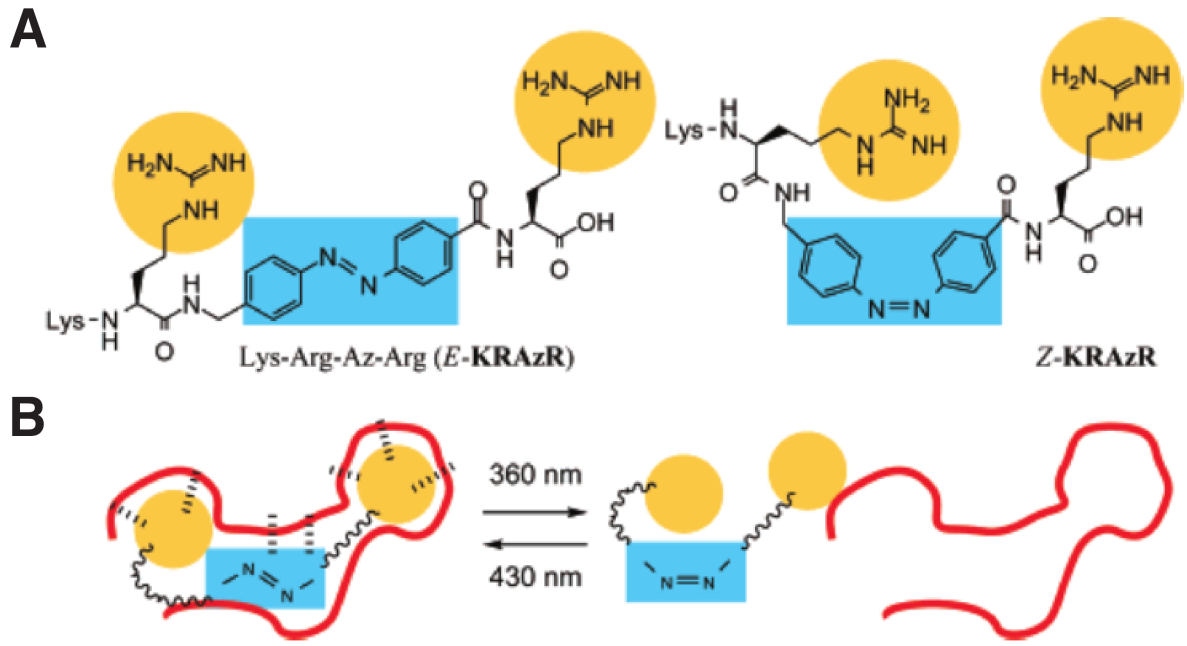


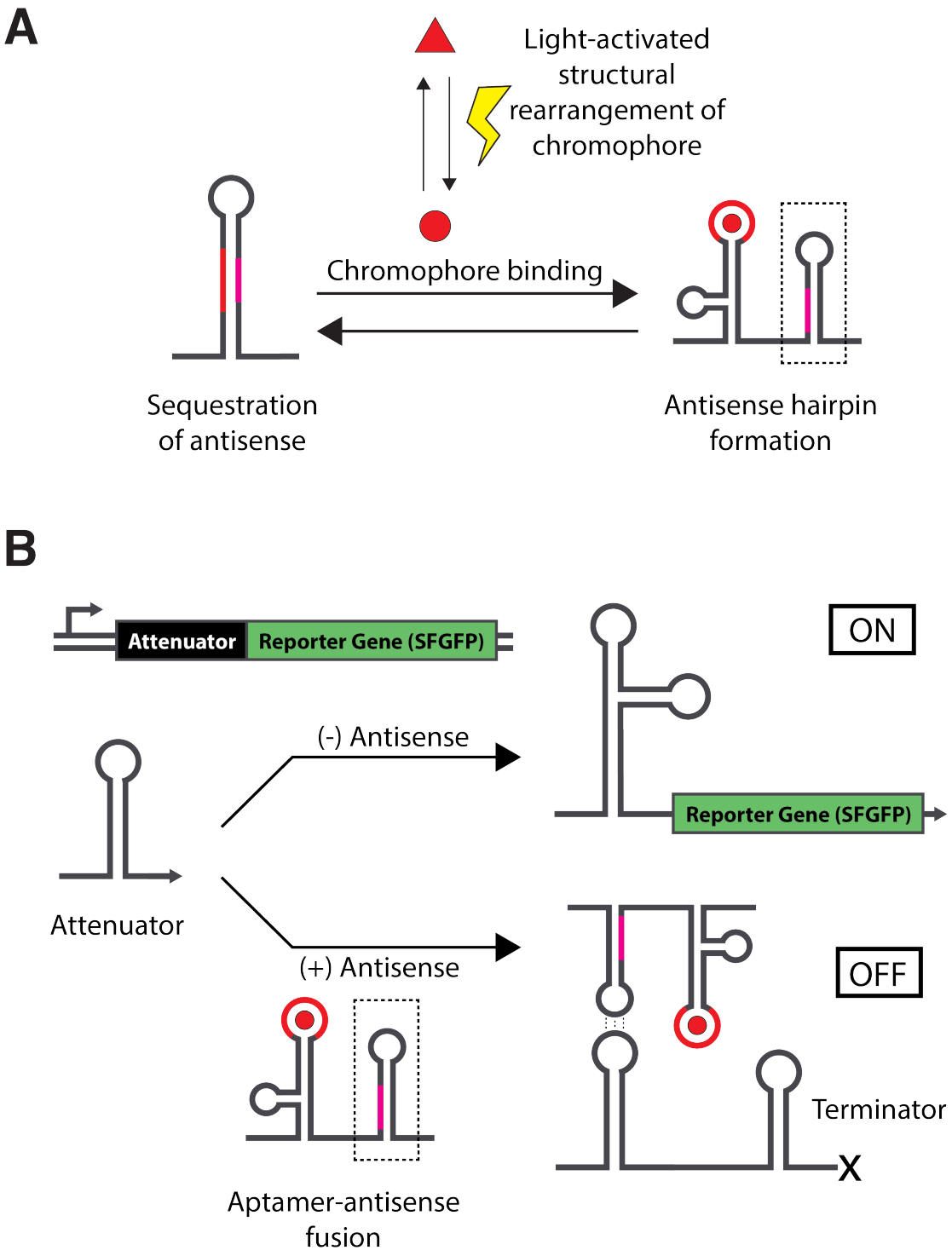
Figure 2.3 Aptamer binding to the light-responsive KRAzR peptide.* **(A)** The KRAzR peptide (Lysine-Arginine-Azobenzene-Arginine) in its *E* and *Z*-isomeric forms, with the azobenzene highlighted in blue and the guanidinium groups that undergo spatial rearrangement highlighted in yellow. **(B)** A schematic of RNA aptamer (red) binding the KRAzR *E*-isomer and releasing upon UV-light catalyzed isomerization to *Z*-KRAzR. *Figure reprinted with permission from Hayashi, G., Hagihara, M., Dohno, C., and Nakatani, K. (2007). Photoregulation of a Peptide-RNA Interaction on a Gold Surface. J. Am. Chem. Soc. *129*, 8678–8679. Copyright (2007) American Chemical Society.

Following a similar principle, Hayashi and coworkers developed a series of RNA aptamers for KRAzR, a short peptide containing a photosensitive azobenzene molecule that reversibly switches between its *E* and *Z*-isomers (Hayashi et al., 2007, 2009) (Figure 2.3A). After using SELEX to isolate aptamers that preferentially bound KRAzR in its *E*-isomer (Figure 2.3B), they established the reversibility of the interaction using SPR to detect aptamer binding to light-switched KRAzR immobilized on a gold surface (Hayashi et al., 2007). Later, they continued with additional cycles of SELEX to identify more aptamers, followed by a doped re-selection of mutants of the best variant. By examining the distribution of sequences in the re-selected pool of mutants, they were able to identify the conserved regions necessary for binding and established a probable structure for the KRAzR aptamer (Hayashi et al., 2009). Their work underlines the flexibility of SELEX to identify RNA aptamers for photo-responsive peptides as well as small molecules.

However, not all photo-sensitive RNAs depend on an aptamer domain: in fact, a photoswitchable hammerhead ribozyme developed by Liang et al. (2013) owes its light-sensitivity to azobenzene-containing oligonucleotides fused to either end of the ribozyme. When the azobenzenes are in their *trans*-form, the oligonucleotides form a duplex that inhibits ribozyme cleavage, but irradiation with UV light triggers azobenzene isomerization to the *cis*-form, allowing the ribozyme to cleave its target. Moreover, the authors demonstrated that visible light could be used to re-set the switch, allowing for reversible photoregulation of ribozyme activity.

Although the use of azobenzene modified nucleotides offered an intriguing way to incorporate photosensitivity into RNAs, I chose to pursue a strategy based on RNA aptamers that bind light-responsive targets. Avoiding the need for artificially synthesized

Figure 2.4 Schematic for proposed light-sensitive RNA transcriptional regulator. **(A)** Rearrangement of the chimeric antisense RNA composed of an aptamer for a light-sensitive chromophore fused to the pT181 antisense. The target chromophore isomerizes between two states (red circle and red triangle) in response to light. When the aptamer binds the chromophore in its binding-competent isomer (red circle), the antisense hairpin (dashed rectangle) forms, but the antisense hairpin is interfered in the absence of chomophore binding. **(B)** Mechanism of action of the pT181 antisense hairpin, whose interaction with the pT181 attenuator causes formation of a terminator hairpin, preventing transcription of the downstream reporter gene (SFGFP) (OFF). In the absence of antisense hairpin RNA, the terminator hairpin does not form, allowing expression of the reporter gene (ON).


A
Light-activated
structural
rearrangement of
chromophore
Chromophore binding
Sequestration
of antisense
Antisense hairpin
formation
B
Attenuator
Reporter Gene (SFGFP)
ON
(-) Antisense
Reporter Gene (SFGFP)
Attenuator
(+) Antisense
OFF
Terminator
X
Aptamer-antisense
fusion

nucleotides left open the possibility of generating light-activated circuitry that could eventually be usable *in vivo.* Our goal was to generate a photosensitive RNA transcriptional regulator and characterize its function and structure using TX-TL assays (see section 1.6 for background) and SHAPE-CE (covered in section 1.4.1). The regulator was to be generated by fusing an RNA aptamer for a light-sensitive ligand to the prototypical pT181 transcriptional attenuator (discussed in Chapter 1.3.1) upstream of the reporter gene super-folder green fluorescent protein (SFGFP) (Figure 2.4). In this chapter, I will outline my attempts toward this goal and summarize my findings along the way.

2.3 Results

2.3.1 Exploring KRAzR-Based RNA Switches

As a potential strategy for generating light-sensitive RNA switches, I explored the possibility of building off previous work (Hayashi et al., 2008) by using the KRAzR aptamer and its azobenzene peptide ligand. While the KRAzR aptamer proved amenable to structural study and was successfully characterized via SHAPE-CE (Figure 2.5) after *in vitro* transcription and purification, acquiring the peptide ligand proved unfeasible, making any approach based on the KRAZr aptamer unviable.

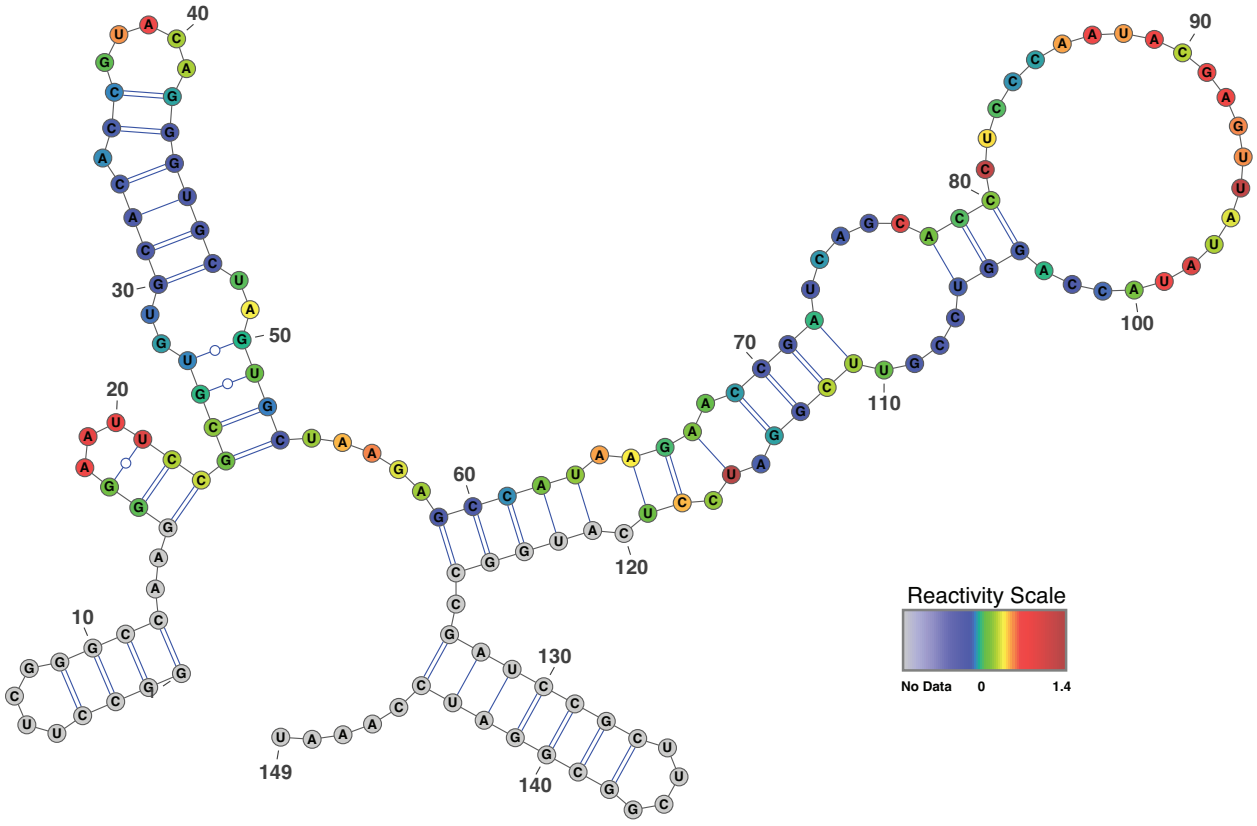


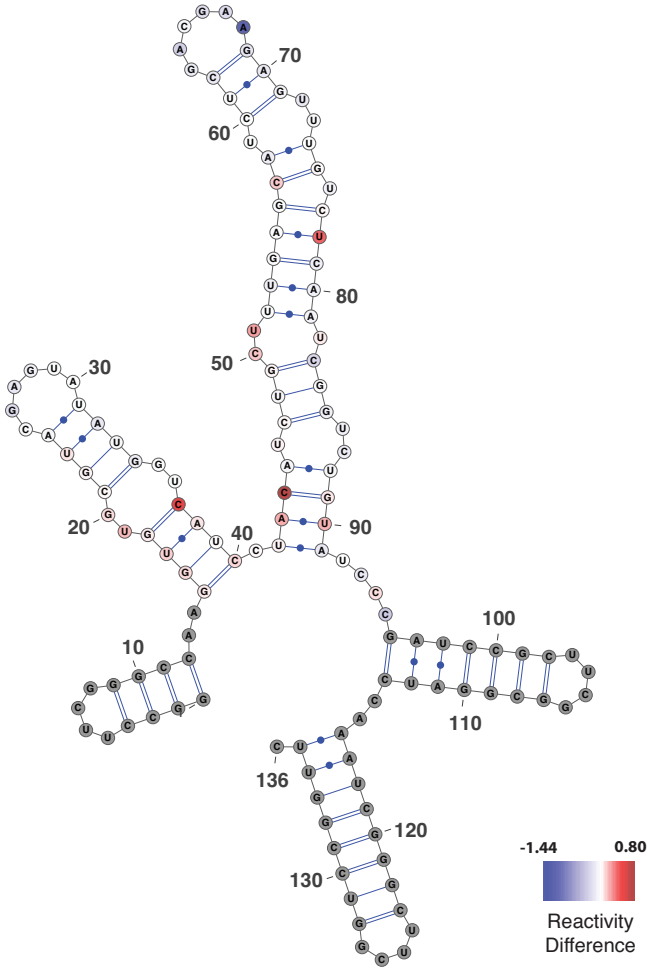
Figure 2.5 Structure of the KRAzR aptamer 58. The folding of the previously reported KRAzR aptamer (Hayashi et al., 2008) is predicted using SHAPE-CE reactivities as constraints in RNAStructure. Highly reactive nucleotides are colored in red, low reactivities nucleotides in blue, and the nucleotides without any data (largely the structure cassettes) in grey.

2.3.2 Dihydropyrene

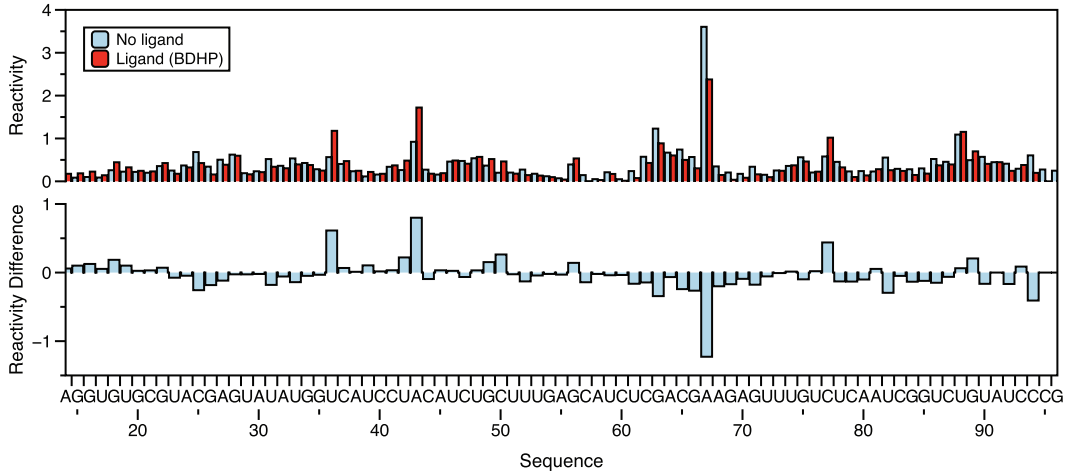
As an alternate mode of encoding light-sensitivity, I investigated using the dihyropyrene aptamer (Clone 8) developed by Lee et al. (2007). The aptamer proved amenable to SHAPE-CE, so I ran probing experiments with and without the target dihydropyrene

Figure 2.6 Structure and reactivity of dihydropyrene aptamers. **(A)** Predicted structure of dihydropyrene atpamer from RNAStructure (Reuter and Mathews, 2010) prediction constrained by SHAPE reactivity data in the absence of ligand. The color map displays the reactivity difference after the addition of ligand, with blue indicating a reactivity drop, red indicating a reactivity increase, white indicating no change, and grey indicating no data. The reactivity pattern suggests ligand-binding at the top stem-loop and possibly the left-hand stem-loop as well. **(B)** SHAPE reactivity and reactivity difference plots showing the reactivity data for the BDHP aptamer. The groupings of reactivity decreases between 25-30 and 65-70 nucleotides correspond to the possible ligand binding stem-loop regions.

A


70
60
80
50
30
20
40
90
10
100
110
136
120
130
-1.44
0.80
Reactivity
Difference

B


No ligand
Ligand (BDHP)
Reactivity
Reactivity Difference
AGGUGUGCGUACGAGUAUAUGGUCAUCCUACAUCUGCUUUGAGCAUCUCGACGAAGAGUUUGUCUCAAUCGGUCUGUAUCCCG
20
30
40
50
60
70
80
90
Sequence

ligand (the PEG-attached variant in its BDHP confirmation). Localized decreases in reactivity confirmed the putative ligand binding site from the original research and suggested the possible involvement of additional distal residues in ligand binding (Figure 2.6). Moreover, absorbance spectra of the dihydropyrene-PEG before and after visible light exposure matched the spectra reported by Lee et al. (2007), confirming the transition between the BDHP (closed) and BCPD (open) forms of the ligand (Figure A.1).

Based on this structural data, I designed and generated a series of minimal dihydropyrene aptamers (Figure 2.7) and established their ability to bind the

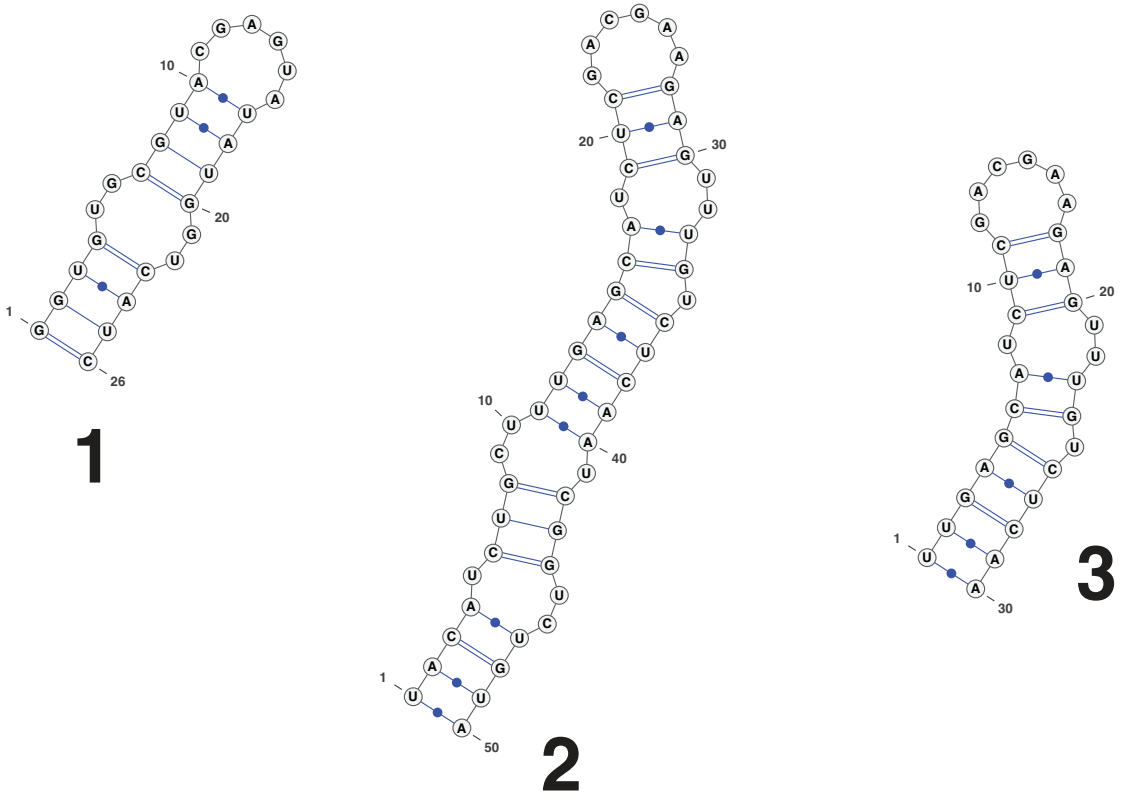


Figure 2.7 Structure of dihydropyrene minimal aptamers. Designed minimal truncations of the dihydropyrene aptamer.

dihydropyrene target using changes in SHAPE reactivity. These minimal aptamers were then fused to the antisense RNA of the pT181 transcriptional regulator following the scheme outlined in Figure 2.4. While functional testing of these fusions in the TX-TL cell-extract system initially gave promising results for the fusion based on minimal hairpin 1 (Figure A.2), appearing to show repression of fluorescence in response to dihydropyrene, multiple further experiments failed to replicate these results. More pT181 fusions were constructed, and while the addition of the aptamer domain did prevent repression just as designed, the addition of ligand failed to restore antisense repression (Figure 2.8).

As an alternate strategy, I followed the riboswitch engineering method laid out by Ceres et al. (2013), in which a variety of aptamer domains were fused to naturally-derived riboswitch expression platforms in order to generate new riboswitches in a modular fashion (see section 1.5). While initial tests verified the function of the constructs reported by Ceres et al., multiple attempts to generate dihydropyrene riboswitches using the same expression platforms proved unsuccessful (Figure A.3), leading to eventual abandonment of this avenue.

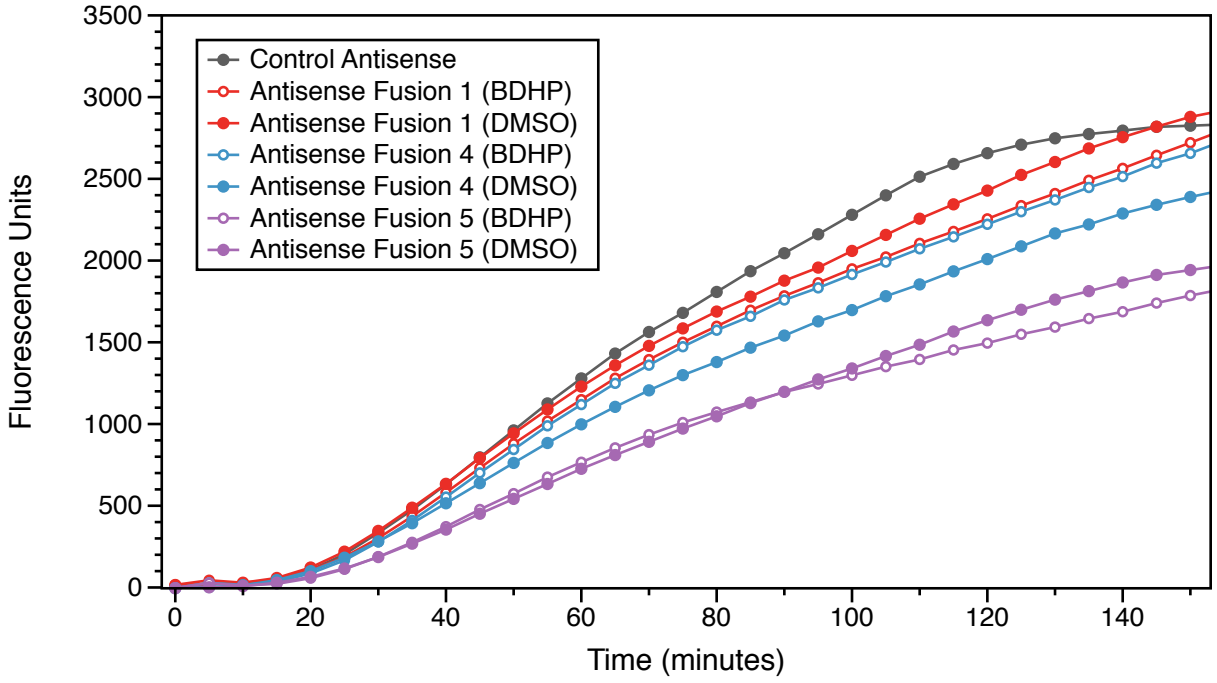


Figure 2.8 TX-TL testing of fusions of the pT181 antisense and dihydropyrene aptamer. Kinetic fluorescence traces were measured in the TX-TL extract system for various antisense plasmid and ligand combinations (BDHP-target ligand, DMSO-no-ligand solvent control) all paired with a pT181 sense plasmid in which the regulator controlled transcription of SFGFP. Functional antisense fusions would be expected to show significantly lower fluorescence in the presence of BDHP (open circles) than DMSO (closed circles).

2.3.3 Testing a spiropyran-sensitive transcriptional switch

Looking to expand off Young and Dieters's success in developing a light-sensitive aptamer (Young and Deiters, 2008), I sought to use their spiropyran-binding aptamer to make a

light-responsive transcriptional regulator. As a starting point, I synthesized the spiropyran aptamer and first established its ability to be characterized via SHAPE-CE. Characterization of spiropyran alone showed that it appeared to transition between its two colorful open forms (see Figure 2.2) in response to changes in pH, and prolonged heat exposure caused the transition between the purple merocyanine form and colorless closed form as reported previously (Young and Deiters, 2008).

As our next step, I cloned a pair of regulators linking the spiropyran aptamer to the pT181 antisense targeting the pT181 transcriptional attenuator upstream of SFGFP (see Figure 2.4). I tested their ON functionality *in vivo*, measuring bulk fluorescence levels in response to the addition of an empty antisense plasmid and the spiropyran aptamer-pT181 fusions. These tests confirmed that the aptamer fusions did not cause termination in the absence of spiropyran (Figure 2.9A), an important checkpoint before embarking on *in vitro* testing using spiropyran. (The costliness of the spiropyran molecule made large-scale culture testing of the with-ligand case impractical.) Moreover, the fusions proved amenable to SHAPE-CE, giving cleanly resolved capillary electropherograms. However, analysis of the reactivity data indicated that the 5' structure cassette was interfering with proper folding of the aptamer (Figure 2.9B), requiring the synthesis of the fusion without the interfering structure cassette.

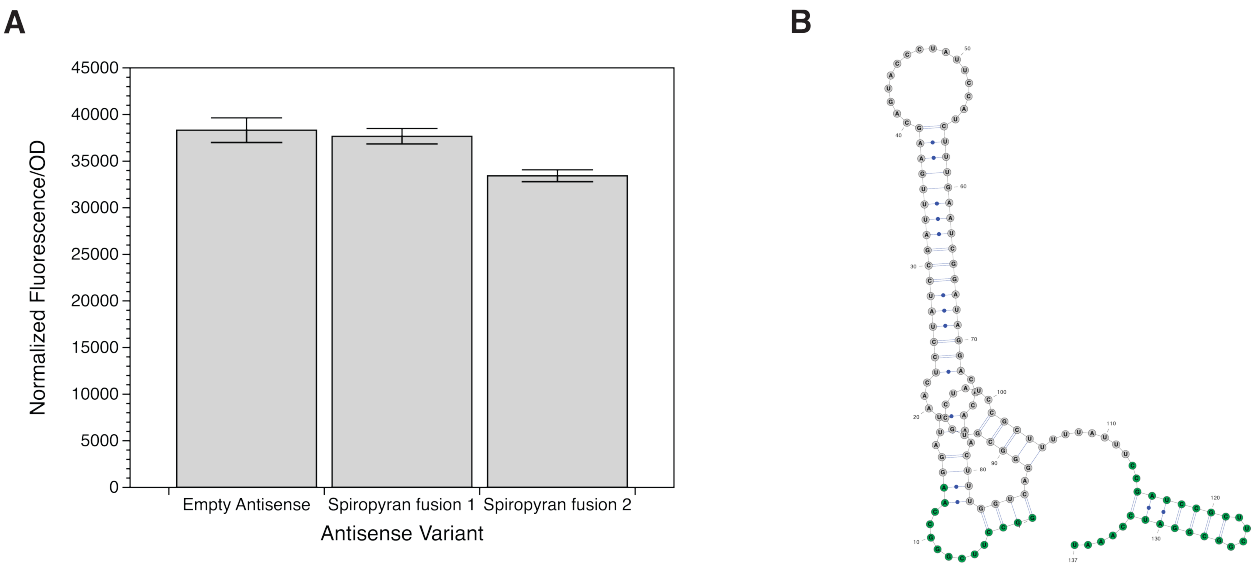


Figure 2.9 Spiropyran aptamer-pT181 fusions. **(A)** *In vivo* functional data showing the normalized fluorescence from the pT181 attenuator-controlled SFGFP output in response to the presence of the spiropyran aptamer-pT181 antisense fusions without spiropyran as compared to an empty control plasmid. As expected, the fusions fail to repress SFGFP fluorescence in the absence of spiropyran. **(B)** Example predicted structure of spiropyran aptamer-pT181 antisense fusion from RNAStructure (Reuter and Mathews, 2010) with prediction constrained by SHAPE reactivity data in the absence of ligand. The structure cassettes (green) appear to be interfering with the folding of the aptamer and antisense rather than folding into self-contained hair-pins as desired.

Attempts to probe the structure of the spiropyran aptamer and fusions in the presence of spiropyran were impeded by an inability to dissolve the spiropyran in the RNA folding buffer at high concentrations. After consulting with the authors of the original research (Young and Deiters, 2008), I addressed this problem by adding 10% DMSO to

solution in order to increase spiropyran solubility. I also used proton NMR to confirm that the spiropyran was adopting the colorless, closed form in solution, although a remaining purple tint hinted at the continued presence of open form.

Although, initial attempts to reverse-transcribe the spiropyran fusions failed to give good extension, optimization of reverse-transcription conditions (temperature and extension time) eventually gave analyzable results. However, SHAPE characterization of the spiropyran aptamer and fusions with and without spiropyran failed to distinguish any reactivity differences associated with ligand binding regardless of the form of the spiropyran (Figure A.6). Contacting Douglas Young helped explain this surprising result—his own experiences indicated that spiropyran behaved differently when immobilized (as in the original selection and SPR binding experiments) as opposed to when free in solution, where it would quickly revert to the closed form. Based on this knowledge and our failure to demonstrate ligand-binding via SHAPE-CE, we chose to discontinue work with the spiropyran aptamer.

2.3.4 Biliverdin aptamer characterization

As an alternate approach to developing a light-sensing RNA activator, I investigated using biliverdin as the relevant ligand. Biliverdin (Figure 2.10) is a green tetrapyrrole pigment that arises naturally in the breakdown of heme (José et al., 2001) and which can undergo changes in conformation in response to light (Bois-Choussy and Barbier, 1980). Moreover, biliverdin has been previously synthesized in *E. coli* at rates reaching 3.3 mg/L (Chen et al., 2012), indicating that *E. coli* can tolerate biliverdin as a possible signaling molecule. Fortuitously, our collaborators in Matthew Levy's lab at Albert Einstein College of Medicine, had already performed ten rounds of SELEX on a randomized RNA library, generating a

pool of sequences enriched for biliverdin-binding RNA aptamers.

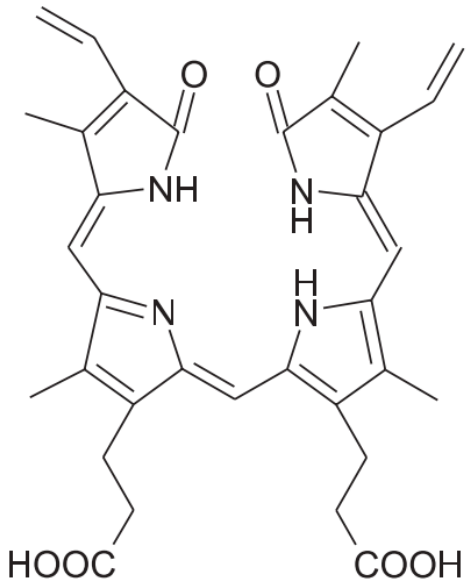
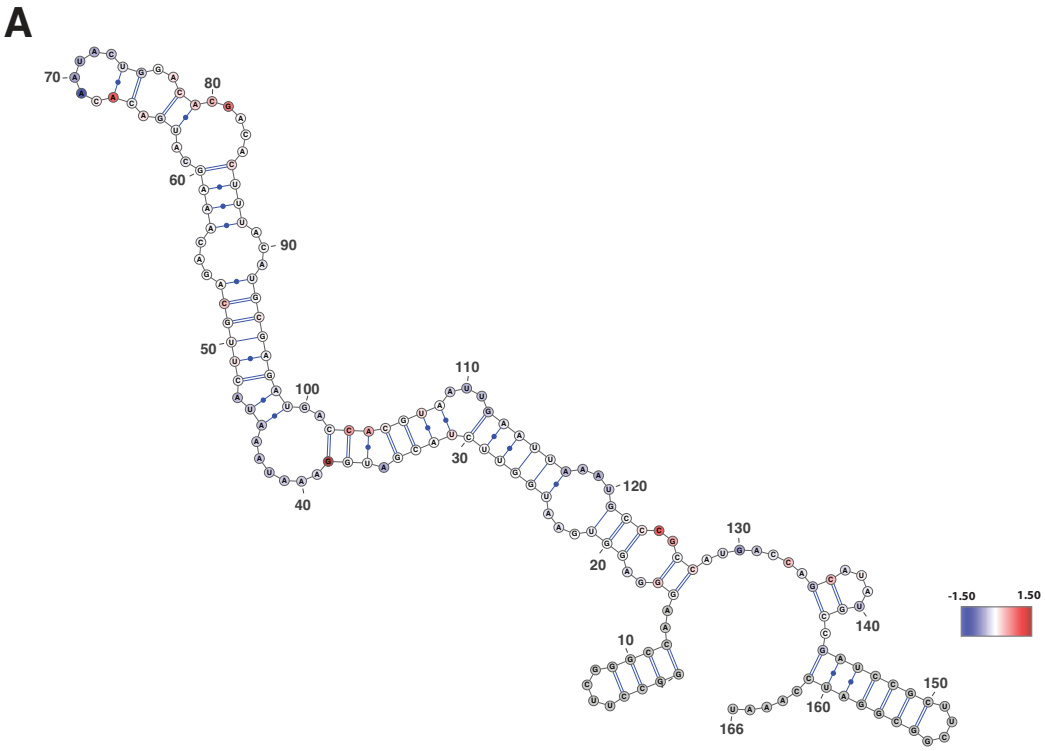


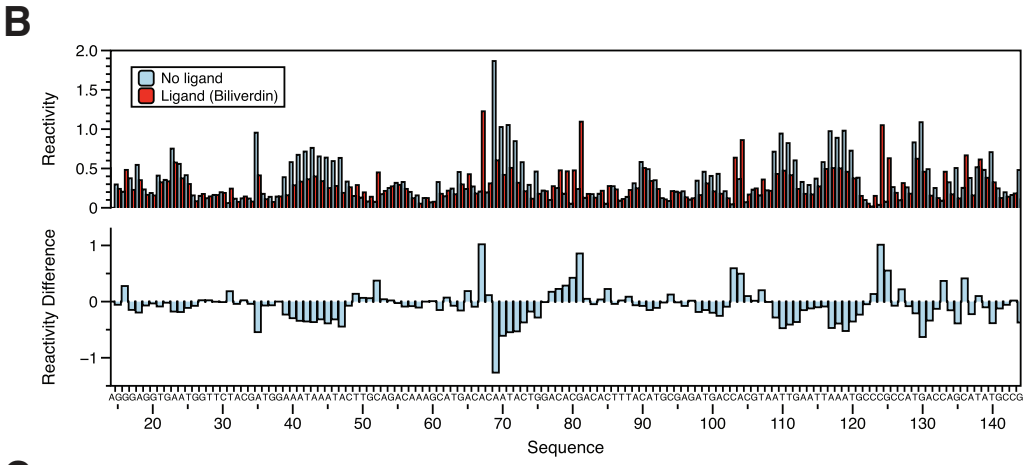
Figure 2.10 The structure of the heme-derived, green tetrapyrrole pigment biliverdin. Reprinted from Kirkby and Adin (2006).

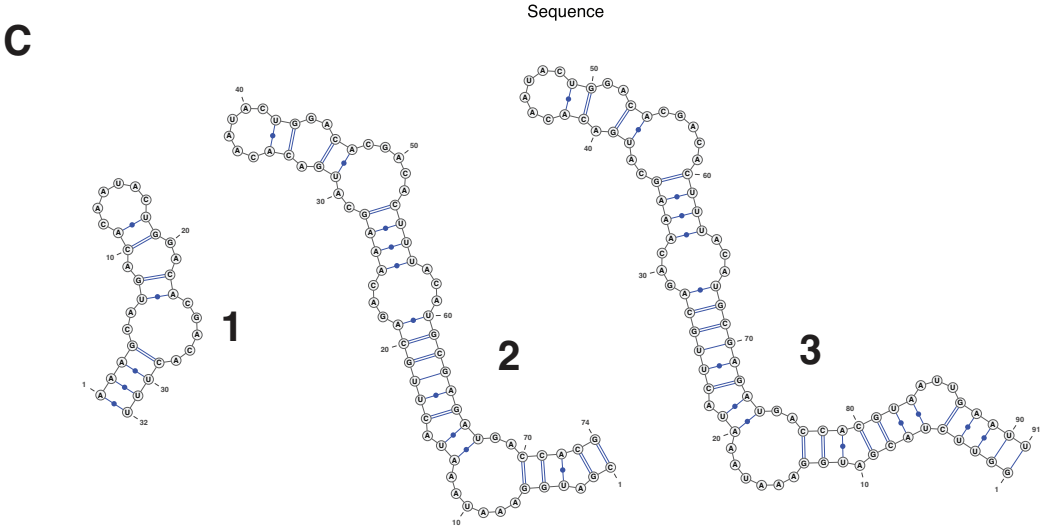
After sequencing this pool of SELEX-generated potential aptamers (Figure A.4), I synthesized a handful of the highest-frequency aptamers via *in vitro* transcription to test for changes in structure in the presence of biliverdin. When the SHAPE-CE reactivities between the ligand and no ligand cases were compared, one of the aptamers tested (aptamer 70), showed particularly striking changes in reactivity patterns, indicating a probable location of biliverdin binding (Figure 2.11A). Based on this data, I designed three different potential minimal aptamers (Figure 2.11B) in order to establish the smallest motif sufficient for ligand binding out of the original 121 nucleotide aptamer. My hope was that

working with a minimal version of the aptamer would make for simpler engineering of our eventual synthetic riboswitch. Of the minimal aptamers tested, the one of intermediate length (minimal aptamer 2, Figure 2.11B) gave the best results: it folded into its intended structure (based on structure prediction constrained by SHAPE reactivities) and exhibited the expected reactivity changes in response to ligand binding. However, further exploration at this point indicated that the conformational switching and relaxation of biliverdin was too fast to be of utility for light-activated riboswitch, making continued in work this direction impractical.

Figure 2.11 Structure and reactivity of biliverdin aptamers. **(A)** Predicted structure of biliverdin aptamer 70 from RNAStructure (Reuter and Mathews, 2010) prediction constrained by SHAPE reactivity data in the absence of ligand. The color map displays the reactivity difference after the addition of ligand, with blue indicating a reactivity drop, red indicating a reactivity increase, white indicating no change, and grey indicating no data. The reactivity pattern suggests ligand-binding at the top stem-loop. **(B)** Reactivity plots showing the SHAPE-CE reactivities with and without ligand, as well as the reactivity differences. **(C)** Designed structures for three minimal truncations of the biliverdin aptamer 70.


A
70
80
60
90
50
100
110
30
40
120
20
130
10
140
150
160
166
-1.50
1.50


B
No ligand
Ligand (Biliverdin)
Reactivity
Reactivity Difference
Sequence


C
1
2
3

2.4 Discussion and conclusions

While the work failed to develop a light-activated RNA switch, the reported research did reaffirm the utility of SHAPE probing for identifying the location of ligand-binding in an RNA aptamer in the case of both the dihydropyrene and biliverdin aptamers. There is no doubt that creating a light-sensitive RNA regulator remains entirely feasible, given that work by Lee et al. (2007) in creating a light-responsive ribozyme has shown that light-sensitive ligand binding can indeed be linked to RNA functionality. However, to engineer such a regulator may require the isolation of new RNA aptamers for a carefully selected target—ideally one with stable, light-switchable conformations, as well as high solubility and bio-compatibility. Regardless, the challenge remains to engineer an RNA that ties a light-based input to an output in gene expression, and it is my hope that this goal is someday achieved.

2.5 Materials and methods

Gene synthesis and plasmid cloning

The KRAZr aptamer templates were synthesized by PCR to create DNA templates for *in vitro* transcription.

The spiropyran aptamer and minimal aptamers were cloned into a pUC57 backbone via Gibson assembly (Gibson et al., 2009) and fused to the pT181 regulator using the inverse polymerase chain reaction (iPCR) (Hemsley et al., 1989) method of site-directed mutagenesis.

The dihydropyrene aptamer was cloned into a colE1 backbone and fused to the pT181 regulator using Gibson assembly.

The biliverdin aptamers were synthesized via PCR and cloned into a Genscript pUC57 backbone provided by Matthew Levy of Albert Einstein College of Medicine using Gibson assembly such that the aptamer sequence was preceded by a T7 promoter and in certain cases followed hepD ribozyme to improve the accuracy of the 3' end.

RNA preparation via in vitro transcription

RNA for structural characterization was transcribed via *in vitro* transcription from its synthetic DNA template including a T7 promoter and flanking structural cassette regions. *In vitro* transcription reactions were generally performed at the 1 mL scale, requiring 20 µL of the resuspended PCR product generated as outlined above. In addition, each transcription reaction contained 40 mM Tris (pH 8.0), 20 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 0.01% Triton, 5 mM mixed NTPs, 3 µL RNAse Inhibitor (Promega), and 10 µL of an in-house His-tagged T7 RNA polymerase preparation. After careful mixing, the reaction was incubated at 37°C for at least 6 hours, until a white organic phosphate precipitate was clearly visible at the bottom of the reaction tube.

RNA purification

The RNA from the *in vitro* transcription was purified by ethanol precipitation followed by polyacrylamide gel extraction. To a given volume of supernatant were added 0.1 volumes of 3 M NaOAc (pH 5.2) and 3 volumes of ice-cold 100% EtOH, and the resulting mixture was chilled for at least 30 minutes at -80 °C before centrifugation at 15,000 rpm at 4 °C for 30 minutes. The supernatant was removed, and the RNA pellet was dissolved in a solution of 85% deionized formamide, 0.5X TBE (Tris/Borate/EDTA) buffer, 50 mM EDTA (pH 8.0),

and 0.01% (w/v) xylene cyanol and bromophenol blue by heating the sample tubes at 95 °C and carefully pipetting up and down to encourage pellet dissolution.

The RNA was run on a large, pre-warmed 8% denaturing polyacrylamide gel for 2.5-3 hrs at 35 W. Bands were visualized using UV shadowing, then cut and transferred to microcentrifuge tubes, crushed with a pipette tip, and passive-eluted into nuclease free water while rotating overnight on a tube rotisserie at 4 °C. The resulting eluted RNA was once again ethanol precipitated by adding 0.1 volumes 4 M NaCl and 3 volumes ice-cold 100% EtOH, followed by 15 minutes' chilling at -80 °C and centrifugation at 15,000 rpm at 4 °C for 30 minutes. Afterwards, the supernatant was removed, the pellet re-centrifuged for 2 minutes, and any remaining supernatant was carefully aspirated away. The RNA was resuspended in TE (Tris/EDTA) buffer and its concentration determined using the Qubit fluorometric quantitation assay. Finally, 150 ng were run on a small denaturing polyacrylamide gel alongside RNA Century Ladder (Ambion) in order to check for purity and correct RNA length. The gel was stained using SYBR Gold and imaged on a ChemiDoc XRS+ (Bio-Rad).

Selective 2'-Hydroxyl acylation Analyzed by Primer Extension-Capillary Electrophoresis (SHAPE-CE) and SHAPE-Seq

SHAPE-Seq experiments were performed following the protocol previously published (Loughrey et al., 2014), and SHAPE-CE experiments were executed as follows:

RNA modification

The modification protocol was based on the original method published by Merino et al. (

(2005), altered to use the SHAPE reagent 1M7 instead of NMIA and to accommodate the addition of ligand during folding:

Ten picomoles of the RNA to be modified were brought up to a total volume of 12 µL in nuclease-free water and heated for 2 minutes at 95 °C, then removed to ice for 1 minute, followed by the addition of 6 uL of 3.3x folding buffer composed of 333 mM HEPES (pH 8.0), 333 mM NaCl, and 33 mM $MgCl_2$. The RNA was refolded for 20 minutes at 37 °C. (In experiments involving the addition of ligand, the RNA volume was decreased to accommodate the ligand volume, and the ligand was added before RNA refolding.) Next, 9 µL of the refolded RNA were added to two pre-prepared tubes containing either 1 µL anhydrous DMSO (unmodified (-) reaction) or 1 µL 65 mM of SHAPE reagent 1M7 dissolved in anhydrous DMSO (modified (+) reaction), and the tubes were both incubated at 37 °C for 1 minute.

To precipitate the RNA, 90 µL of nuclease-free water, 5 µL of 3 M NaOAc, 1 µL of glycogen (to aid in visualizing RNA pellet), and 400 µL ice-cold 100% EtOH were added to each tube, and the tubes were then incubated at -80 °C for at least 30 minutes. Next, the tubes were centrifuged at 15,000 rpm at 4 °C for 30 minutes, followed by removal of the supernatant and 2 additional minutes of centrifugation. Any final traces of supernatant were aspirated, and the RNA pellets were dissolved in 10 µL 0.5x TE buffer and stored at -20 °C.

Reverse transcription

For each RNA to be characterized, 20 pmol were dissolved in 18 µL of nuclease-free water and split into two tubes with 9 µL each, corresponding to the ddA and ddT sequencing

reactions. To each of these sequencing tubes and the (+) and (-) reactions from the modification step were added 3 μL of a fluorescently labeled reverse transcription primer complementary to its target, with fluorescent labeling corresponding to the particular reaction: 0.3 μM of primers labeled with FAM (ddT sequencing reaction), VIC ((+) modified RNA reaction), or NED ((-) un-modified RNA reaction), or 0.6 μM of primer labeled with PET (ddA sequencing reaction). The reactions were heated on a thermal cycler at 95 °C for two minutes and then 65 °C for 5 minutes before removal to ice. Next, 1 μL ddA and ddT was added to the appropriate tubes, and 1 μL 0.5x SSIII (Invitrogen) and 6 μL Master Mix* were added to every tube. The tubes were then heated at 45 °C for 1 minute, 52 °C for 25 minute, 65 °C for 5 minutes, and finally cooled to 4 °C.

*Master Mix composition: 4 volumes of 5x SSIII First Strand Buffer (Invitrogen), 1 volume 0.1 M DTT, and 1 volume 10 mM mixed dNTPs (10 mM each dATP, dCTP, dTTP, and dITP).

The reverse-transcribed DNA from each sample was precipitated together, with the sequencing and (+) and (-) reactions for each sample combined along with 240 μL ice-cold 100% EtOH into a non-stick 2 mL micro-centrifuge tube. The tubes were chilled at -80 °C for at least 15 minutes before centrifugation at 4 °C and 15,000 rpm. The supernatant was removed, and the precipitated DNA pellets were washed with 800 μL ice-cold 70% EtOH and centrifuged for a further 2 minutes. The supernatant was pipetted off again, the pellets re-spun for a final 2 minutes, and any remaining supernatant was aspirated before the pellets were dried for 10 minutes on a rotary evaporator set on its lowest setting. The dried pellets were resuspended in 10 μL deionized formamide before submission for fragment analysis on an ABI 3730xl capillary electrophoresis machine. Data were aligned and analyzed using SHAPEFinder (Vasa et al., 2008), and reactivities were normalized

using the 2%/8% rule (Lucks et al., 2011).

TX-TL testing

TX-TL extract was prepared following the previously reported protocol (Shin and Noireaux, 2010; Sun et al., 2013). Plasmids to be tested were grown in NEB Turbo cells overnight and purified following the manufacturer protocol for the QIAfilter Plasmid Midi Kit (Qiagen), deviating from the protocol only in that the plasmid pellets were dissolved in MilliQ-purified water in place of buffer. The TX-TL experiments were performed following the methods laid out by Takahashi et al. (2014) with temperature set at either 30 ° C or 37 ° C and fluorescent readings taking place every 5 minutes.

In vivo functional testing

All *in vivo* bulk fluorescence experiments were performed in *Escherichia coli* (*E. coli*) strain TG1 *(F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB))*. For each experiment, pairs of sense target plasmid and antisense plasmid (or a no-antisense control plasmid) were transformed into chemically competent *E. coli* TG1 cells, plated on Difco LB + Agar plates with 100 mg/mL carbenicillin and 34 mg/ml chloramphenicol, and incubated overnight at 37 °C for approximately 17 hours. For every experiment, the same procedure was repeated for the pair of empty autofluorescence control plasmids, JBL001 and JBL002, which did not express either antisense or sense target RNAs and were used to determine background cellular autofluorescence. After overnight incubation, the plates were removed from the incubator and kept at room temperature for approximately 7 hours. Three separate colonies were picked for each

condition or control, and each was used to inoculate 300 µL of LB (Luria-Bertani) media containing 100 mg/mL carbenicillin and 34 mg/mL chloramphenicol in a 2 mL 96-well block (Costar 3960). The block was covered with a breathable seal (Aeraseal BS-25) and incubated at 37 °C while shaking at a speed of 1,000 rpm in a Labnet Vortemp 56 bench-top shaker for 18-19 hours overnight. From this overnight culture 4 µL were taken and used to inoculate 196 µL of M9 minimal media (1 x M9 minimal salts, 1 mM thiamine hydrochloride, 0.4 % glycerol, 0.2 % casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) containing 100 mg/mL carbenicillin and 34 mg/ml chloramphenicol. The cultures, alongside three wells of an M9 only control, were grown in a 2 mL 96-well block under the same conditions as the overnight culture for ~4 hours. Fifty µL of this culture were diluted 1:1 with phosphate buffered saline (PBS) in a black-welled clear-bottomed 96-well plate (Costar 3231). The diluted cultures' optical density (OD) at 600 nm and fluorescence (485 nm excitation, 520 nm emission) were then measured with a Biotek Synergy H1 plate reader.

Data analysis

After subtracting the average absorbance and fluorescence of the three wells containing the media blank from the measured sample values, the ratio of fluorescence to OD was calculated for each well, and the triplicate values were averaged for each set of samples after subtraction of the average fluorescence to OD ratio for the background cellular autofluorescence control.

2.6 Acknowledgements

I would like to extend my deepest gratitude to the undergraduate researchers Sitara Sankar and Rebecca Chew, who contributed greatly to this project and performed many of the key experiments. I am also grateful to our collaborators Matthew Levy and Simon Trevino of Albert Einstein College of Medicine for the biliverdin aptamer library and their valuable assistance. Additional thanks go to Alexander Deiters and Doug Young for their kind help in answering my questions regarding spiropyran and to Dipankar Sen for providing the dihydropyrene compound for our studies. I would also like to thank Ivan Keresztes, director of the Cornell University NMR facility, for lending his expertise in interpreting the spiropyran NMR structure.

2.7 Bibliography

Airan, R.D., Thompson, K.R., Fenno, L.E., Bernstein, H., and Deisseroth, K. (2009). Temporally precise in vivo control of intracellular signalling. Nature *458*, 1025–1029.

Asanuma, H., Shirasuka, K., Yoshida, T., Takarada, T., Liang, X., and Komiyama, M. (2001). Spiropyran as a Regulator of DNA Hybridization with Reversed Switching Mode to That of Azobenzene. Chem. Lett. 108–109.

Bois-Choussy, M., and Barbier, M. (1980). Photochemistry of Biliverdin IX δ as a Model for the Study of the Photoproducts from Natural Biliverdin IX ϒ (Pterobilin). Helv. Chim. Acta *63*, 1098–1109.

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci. *8*, 1263–

1268.

Cambridge, S.B., Geissler, D., Keller, S., and Cürten, B. (2006). A caged doxycycline analogue for photoactivated gene expression. Angew. Chemie *45*, 2229–2231.

Cao, J., Arha, M., Sudrik, C., Bugaj, L.J., Schaffer, D. V, and Kane, R.S. (2013). Light-inducible activation of target mRNA translation in mammalian cells. Chem. Commun. *49*, 8338–8340.

Ceres, P., Trausch, J.J., and Batey, R.T. (2013). Engineering modular "ON" RNA switches using biological components. Nucleic Acids Res. *41*, 10449–10461.

Chen, D., Brown, J.D., Kawasaki, Y., Bommer, J., and Takemoto, J.Y. (2012). Scalable production of biliverdin IXα by Escherichia coli. BMC Biotechnol. *12*, 89.

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

Gardner, L., and Deiters, A. (2012). Light-controlled synthetic gene circuits. Curr. Opin. Chem. Biol. *16*, 292–299.

Gautier, A., Deiters, A., and Chin, J.W. (2011). Light-activated kinases enable temporal dissection of signaling networks in living cells. J. Am. Chem. Soc. *133*, 2124–2127.

Gibson, D.G., Young, L., Chuang, R., Venter, J.C., Hutchison, C. a, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Hayashi, G., Hagihara, M., Dohno, C., and Nakatani, K. (2007). Photoregulation of a Peptide-RNA Interaction on a Gold Surface. J. Am. Chem. Soc. *129*, 8678–8679.

Hayashi, G., Hagihara, M., and Nakatani, K. (2008). RNA aptamers that reversibly bind

photoresponsive azobenzene-containing peptides. Chembiochem *15*, 424–432.

Hayashi, G., Hagihara, M., Nakatani, K., Dohno, C., and Nakatani, K. (2009). RNA aptamers that reversibly bind photoresponsive azobenzene-containing peptides. Chemistry *15*, 424–432.

Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G., Galas, D.J., and Angeles, L. (1989). A simple method for site-directed mutagenesis using the pdymerase chain reaction. Nucleic Acids Res. *17*, 6545–6551.

Jäschke, A. (2012). Genetically encoded RNA photoswitches as tools for the control of gene expression. FEBS Lett. *586*, 2106–2111.

José, P., Pereira, B., Macedo-Ribeiro, S., Párraga, A., Pérez-Luque, R., Darcy, K., Mantle, T.J., and Coll, M. (2001). Structure of human biliverdin IXβ reductase, an early fetal bilirubin IXβ producing enzyme. Nat. Struct. Mol. Biol. *8*, 215–220.

Kirkby, K., and Adin, C. (2006). Products of heme oxygenase and their potential therapeutic applications. Am. J. Physiol. *290*, F563–F571.

Kralj, J.M., Douglass, A.D., Hochbaum, D.R., Maclaurin, D., and Cohen, A.E. (2012). Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. Nat. Methods *9*, 90–95.

Lee, H.-W., Robinson, S.G., Bandyopadhyay, S., Mitchell, R.H., and Sen, D. (2007a). Reversible photo-regulation of a hammerhead ribozyme using a diffusible effector. J. Mol. Biol. *371*, 1163–1173.

Lee, H.-W., Robinson, S.G., Bandyopadhyay, S., Mitchell, R.H., and Sen, D. (2007b). Reversible photo-regulation of a hammerhead ribozyme using a diffusible effector. J. Mol. Biol. *371*,

1163–1173.

Lemke, E.A., Summerer, D., Geierstanger, B.H., Brittain, S.M., and Schultz, P.G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. Nat. Chem. Biol. *3*, 769–772.

Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M., et al. (2005). Engineering Escherichia coli to see light. Nature *438*, 441.

Levskaya, A., Weiner, O.D., Lim, W.A., and Voigt, C.A. (2009). Spatiotemporal control of cell signalling using a light-switchable protein interaction. Nature *461*, 997–1001.

Liang, X., Zhou, M., Kato, K., and Asanuma, H. (2013). Photoswitch nucleic acid catalytic activity by regulating topological structure with a universal supraphotoswitch. ACS Synth. Biol. *2*, 194–202.

Loughrey, D., Watters, K.E., Settle, A.H., and Lucks, J.B. (2014). SHAPE-Seq 2.0 : systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. *42*, 1–10.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005). RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J. Am. Chem. Soc. *127*, 4223–4231.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Ryu, M., and Gomelsky, M. (2014). Near-infrared Light Responsive Synthetic c - di-GMP Module for Optogenetic Applications. ACS Synth. Biol. *3*, 802–810.

Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. J. Biol. Eng. *4*, 8.

Song, C., and Knöpfel, T. (2015). Optogenetics enlightens neuroscience drug discovery. Nat. Rev. Drug Discov. *15*, 97–109.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J. Vis. Exp. *79*, e50762.

Tabor, J.J., Salis, H.M., Simpson, Z.B., Chevalier, A.A., Levskaya, A., Marcotte, E.M., Voigt, C.A., and Ellington, A.D. (2009). A synthetic genetic edge detection program. Cell *137*, 1272–1281.

Takahashi, M.K., Chappell, J., Hayes, C.A., Sun, Z.Z., Kim, J., Singhal, V., Spring, K.J., Al-Khabouri, S., Fall, C.P., Noireaux, V., et al. (2015). Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. ACS Synth. Biol. *4*, 503–515.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510.

Vasa, S.M., Guex, N., Wilkinson, K. a, Weeks, K.M., and Giddings, M.C. (2008). ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity

information resolved by capillary electrophoresis. RNA *14*, 1979–1990.

Ye, H., Baba, M.D., Peng, R., and Fussenegger, M. (2011). A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. Science *332*, 1565–1568.

Young, D.D., and Deiters, A. (2007). Photochemical activation of protein expression in bacterial cells. Angew. Chemie *46*, 4290–4292.

Young, D.D., and Deiters, A. (2008). Light-Regulated RNA–Small Molecule Interactions. Chembiochem *9*, 1225–1228.

Zhang, F., Vierock, J., Yizhar, O., Fenno, L.E., Tsunoda, S., Kianianmomeni, A., Prigge, M., Berndt, A., Cushman, J., Polle, J., et al. (2011). The microbial opsin family of optogenetic tools. Cell *147*, 1446–1457.

CHAPTER 3

IMPROVING FOLD ACTIVATION OF SMALL TRANSCRIPTION ACTIVATING RNAS (STARS) WITH RATIONAL RNA ENGINEERING STRATEGIES.[1]

3.1 Abstract

Regulatory RNAs have become integral components of the synthetic biology and bioengineering toolbox for controlling gene expression. We recently expanded this toolbox by creating small transcription activating RNAs (STARs) that act by disrupting the formation of a target transcriptional terminator hairpin placed upstream of a gene. While STARs are a promising addition to the repertoire of RNA regulators, much work remains to be done to optimize the fold activation of these systems. Here we apply rational RNA engineering strategies to improve the fold activation of two STAR regulators. We demonstrate that a combination of promoter strength tuning and multiple RNA engineering strategies can improve fold activation from 5.4-fold to 13.4-fold for a STAR regulator derived from the pbuE riboswitch terminator. We then validate the generality of our approach and show that these same strategies improve fold activation from 2.1-fold to 14.6-fold for an unrelated STAR regulator, opening the door to creating a range of additional STARs to use in a broad array of biotechnologies. We also establish that the optimizations preserve the orthogonality of these STARs between themselves and a set of RNA transcriptional repressors, enabling these optimized STARs to be used in sophisticated circuits.

[1] This work was originally published in *Biotechnology and Bioengineering* and has been reproduced here under the author use and publication permission policy of Wiley. Meyer, S., Chappell, J., Sankar, S., Chew, R., Lucks, J. B. (2016). Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. Biotechnology and Bioengineering *113*, 216-225.

3.2 Introduction

Natural and engineered RNA regulators have become powerful components of our toolbox for precisely regulating gene expression (Chappell et al., 2013). This is in large part due to advances in our understanding of RNA biology that have uncovered a vast range of regulatory functions performed by naturally occurring RNAs (Cech and Steitz, 2014; Chappell et al., 2013). Many of these functions involve the regulation of the fundamental processes of gene expression, including mRNA degradation (Collins et al., 2007; Filipowicz et al., 2008; Storz et al., 2011), translation (Gottesman and Storz, 2011; Nou and Kadner, 2000; Winkler et al., 2002), and transcription elongation (Brantl and Wagner, 2000). Recent work has further revealed how these functions are intimately linked to the structure of the regulatory RNAs and the structural rearrangements they induce in their targets (DebRoy et al., 2014). This in turn has enabled significant advances in design approaches that use computational RNA structure prediction algorithms to design synthetic RNAs that adopt specific conformations to perform their regulatory function (Green et al., 2014; Rodrigo et al., 2012). RNA regulators thus represent a versatile and designable platform for controlling gene expression and have been used in a number of recent applications, including the creation of synthetic RNA regulatory gene expression switches (Ceres et al., 2013a; Lynch et al., 2007; Wachsmuth et al., 2013), RNA–only logic gates and circuits (Chappell et al., 2015; Lucks et al., 2011; Xie et al., 2011), RNA transcriptional networks (Bhadra and Ellington, 2014; Lucks et al., 2011; Takahashi et al., 2014), and RNA-based diagnostics (Pardee et al., 2014).

For bacterial systems, small RNAs (sRNAs) have proven to be particularly well suited to engineering approaches that optimize and alter their function. Although sRNAs

can act through a wide variety of mechanisms (Gottesman and Storz, 2011), a common mode of sRNA regulation relies on Watson-Crick base pairing between a sense target RNA, usually located upstream of the gene to be controlled, and a *trans*-acting antisense sRNA. By itself, the sense target RNA can fold into structures that block or allow gene expression – for example, by occluding a ribosome binding site, in the case of translation regulation, or forming an intrinsic terminator hairpin, in the case of transcription regulation. Interaction between the sense target RNA and antisense sRNA can then cause structural rearrangement, ultimately controlling the expression of the gene. Years of research have uncovered design principles for these mechanisms, enabling engineers to create a wide array of sRNA regulators, including translational repressors (Mutalik et al., 2012; Na et al., 2013) translational activators (Green et al., 2014; Isaacs et al., 2004; Rodrigo et al., 2012), and transcriptional repressors (Lucks et al., 2011; Qi et al., 2012a; Takahashi and Lucks, 2013).

Despite the versatility of engineered sRNA regulators, until recently there were no known natural or synthetic examples of sRNAs that could activate transcription (Chappell et al., 2013). To address this gap, we created small transcription activating RNAs (STARs) (Chappell et al., 2015) (Figure 3.1A). In the STAR mechanism, the sense target region contains an intrinsic terminator hairpin, which terminates transcription in the OFF state, preventing read-through of the downstream gene. The STAR antisense contains a specific anti-terminator sequence that is designed to bind to the 5' stem of the terminator in *trans* to prevent terminator formation and allow transcriptional read-through in the ON state. This mechanistic design strategy was applied to target a range of intrinsic terminators derived from natural sources, from the pbuE riboswitch to the pT181 plasmid copy control

element, ultimately creating five different STARs that displayed a range of transcriptional activation from 3-fold to 94-fold (Chappell et al., 2015). In addition, orthogonality between these STARs and a preexisting library of RNA transcriptional repressors (Takahashi and Lucks, 2013) allowed the construction of two previously unattainable RNA-only logic gates (Chappell et al., 2015), demonstrating the potential of STARs for engineering sophisticated RNA genetic circuitry.

STARs thus represent a powerful expansion of the RNA engineering toolbox for precisely regulating gene expression and creating synthetic genetic networks. However, much work remains to be done to broadly optimize the fold activation of these new regulators. In particular, only two of the originally designed STARs showed levels of activation greater than 10-fold, thereby limiting the number of STARs useful in applications that require large differences between ON and OFF expression states. Here we remedy this by applying several sRNA engineering methods (Carrier and Keasling, 1997; Sakai et al., 2013) and gene expression optimization strategies to optimize the overall fold activation of weak STAR regulators. Specifically, we used a combination of promoter strength tuning and STAR antisense RNA stabilization and engineering strategies to improve fold activation from 5.4-fold (±2.2) to 13.4-fold (±3.8) for the pbuE STAR regulator. To confirm that our approach could be generalized, we then applied these strategies to the unrelated prgX STAR (derived from a conjugation control system terminator) to improve its fold activation from 2.1-fold (±0.4) to 14.6-fold (±3.7). This process also yielded multiple STAR variants for both systems with intermediate fold activation levels, showing that this strategy can be used to fine-tune STAR performance. Finally, although the optimization strategies required the addition of a significant amount of extra RNA sequence and structure, we demonstrated

that orthogonality was preserved between the optimized STARs and a set of RNA transcriptional repressors. These optimization strategies open the door for creating a range of additional STARs for use in a broad array of biotechnologies.

3.3 Materials and Methods

3.3.1 Plasmid construction and cloning

STAR-mediated gene expression was tested with a two-plasmid system. Plasmids were constructed so that the sequence encoding each sense target RNA was placed downstream of a constitutive promoter and upstream of the coding sequence for the superfolder green fluorescent protein (SFGFP) reporter (Pédelacq et al., 2006), complete with its own ribosome binding site (Figure B.1). Separate plasmids were constructed for STAR antisense expression, with the sequence encoding the STAR preceded by a constitutive promoter and followed by the t500 transcriptional terminator (Yarnell and Roberts, 1999) (Figure B.1). For experiments in which the STAR antisense was absent, a control plasmid was constructed containing the constitutive promoter followed directly by a transcriptional terminator (rrnB terminator 'TrrnB') (Figure B.1).

All plasmids and sequences used in this study are enumerated in Table B.1. All sense target plasmids included the p15A origin and a gene for chloramphenicol resistance, while all STAR antisense plasmids contained the ColE1 origin and encoded a gene for carbenicillin resistance. All plasmids were either previously reported or constructed from previously reported plasmids (Chappell et al., 2015; Lucks et al., 2011; Takahashi and Lucks, 2013) using inverse polymerase chain reaction (iPCR) to make substitutions and/or insertions. The inserted sRNA scaffold and stability hairpin sequences were derived from

previous work by Sakai et al. (2013) and Carrier et al. (1999), respectively. Sequence-verified stocks of the plasmids were used for all experiments.

3.3.2 Strains, media, and in vivo bulk fluorescence experiments

All bulk fluorescence experiments were performed in *Escherichia coli* (*E. coli*) strain TG1 *(F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB))* with three independent replicates, except for the *hfq* knockout experiments (Figure 3.3D), which were performed in *E. coli* strain BW25113 *(F^-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM^-, rph-1, DE(rhaD-rhaB)568, hsdR514)* and the BW25113 Δ*hfq* variant from the Keio collection (Baba et al., 2006). For each independent replicate, pairs of sense target plasmid and STAR antisense plasmid (or no-antisense control plasmid JBL002) were transformed into chemically competent *E. coli* TG1 cells, plated on Difco LB + Agar plates with 100 mg/mL carbenicillin and 34 mg/ml chloramphenicol, and incubated overnight at 37 °C for approximately 17 hours. For every experiment, the same procedure was repeated for the pair of empty autofluorescence control plasmids, JBL001 and JBL002, which did not express either STAR or sense target RNAs and were used to determine background cellular autofluorescence. After overnight incubation, the plates were removed from the incubator and kept at room temperature for approximately 7 hours. Three separate colonies were picked for each condition/control, and each was used to inoculate 300 µL of LB media with 100 mg/mL carbenicillin and 34 mg/mL chloramphenicol in a 2 mL 96-well block (Costar 3960). In the case of the pbuE variants in Figure 3.4A, 9 colonies were picked in order to guarantee enough passed the growth requirements described below. The block was covered with a breathable seal

(Aeraseal BS-25) and incubated at 37 °C while shaking at a speed of 1,000 rpm in a Labnet Vortemp 56 bench-top shaker for 18-19 hours overnight. From this overnight culture 4 µL were taken and used to inoculate 196 µL of M9 minimal media (1 x M9 minimal salts, 1 mM thiamine hydrochloride, 0.4 % glycerol, 0.2 % casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) containing 100 mg/mL carbenicillin and 34 mg/mL chloramphenicol. The cultures, alongside three wells of an M9 only control, were grown in a 2 mL 96-well block under the same conditions as the overnight culture, until the majority of the OD values exceeded 0.07, which took 5 to 7 hours. Fifty µLs of this culture were diluted 1:1 with phosphate buffered saline (PBS) in a black-welled clear-bottomed 96-well plate (Costar 3231). The diluted cultures' optical density (OD) at 600 nm and fluorescence (485 nm excitation, 520 nm emission) were then measured with a Biotek Synergy H1 plate reader.

3.3.3 Data analysis for bulk fluorescence experiments

Each experiment included two sets of controls: three wells of a media blank (M9) and three wells inoculated from separate colonies of the autofluorescence control *E. coli* cells lacking SFGFP but harboring plasmids with the same backbones and resistances as all sense target and STAR antisense plasmids (transformed control JBL001 and JBL002). All fluorescence and OD values for each colony were initially corrected by subtracting the corresponding values from the average of the three media blanks. The ratio of fluorescence units over OD (Fluorescence/OD) was then calculated for each well and corrected for background fluorescence by subtracting the average Fluorescence/OD for the autofluorescence control cells without SFGFP. For each STAR-target pair, three independent colonies were characterized from three independent transformations (9 colonies total). Data were

discarded for colonies that showed low growth (OD<0.07), although this requirement was relaxed for the orthogonality grid in Figure 3.5 in order to account for the different growth rates of the tested variants. Averages and standard deviations (depicted by error bars) of Fluorescence/OD were calculated over the repeat experiments. Fold activation was calculated by dividing the average corrected Fluorescence/OD for a STAR-target pair (ON) by the average corrected Fluorescence/OD for the same sense target plasmid paired with the control no STAR antisense plasmid (JBL002) (OFF). Fold activation error was calculated using standard error propagation formulas based on the standard deviations of the average corrected Fluorescence/OD values. In calculating fold repression (Figure 3.5), the negative reciprocal was taken to give the fold repression, i.e., 0.20 became −5-fold repression (Chappell et al., 2015). Statistical significance calculations (two-sided T-tests) were performed on the individual colony Fluorescence/OD ON levels normalized by the average Fluorescence/OD OFF levels.

3.4 Results and Discussion

3.4.1 The pbuE STAR as a case study for optimization of fold activation

As a starting point for exploring strategies to increase fold activation, we chose to focus on a STAR-target system (Chappell et al., 2015) derived from the intrinsic terminator of the pbuE riboswitch (Ceres et al., 2013b). As illustrated in Figure 3.1A, the pbuE STAR is designed to be fully complementary to the 5' half of the pbuE intrinsic terminator present in the target RNA. Interaction of the STAR with its target RNA thus prevents the formation of the terminator hairpin, enabling transcription elongation into the downstream reporter gene, superfolder green fluorescent protein (SFGFP). To determine fold activation, gene

expression was characterized by measuring fluorescence normalized by optical density (Fluorescence/OD) for cultures of *E. coli* cells co-transformed with two plasmids: one plasmid encoding the pbuE sense target fused to the downstream SFGFP coding sequence and the other plasmid encoding either the STAR antisense (ON state) or an empty backbone control (OFF state) (absence of STAR antisense case). Activation was then calculated as a ratio of the ON/OFF Fluorescence/OD values (see Materials and Methods). Characterization of the pbuE STAR in this study reconfirmed its low fold activation, previously reported by Chappell et al. (2015), with an observed activation of 5.4-fold (±2.2) in the presence of the STAR antisense compared to when only the target RNA was expressed (Figure 3.1B). This result indicated that there was ample room for improvement in fold activation compared to the 94-fold (±26) activation shown by the best STAR activator reported previously (Chappell et al., 2015). Furthermore, previous work on applying RNA sequence optimization strategies did not improve the low fold activation of the pbuE STAR system (Chappell et al., 2015), motivating us to pursue a suite of alternative strategies discussed below.

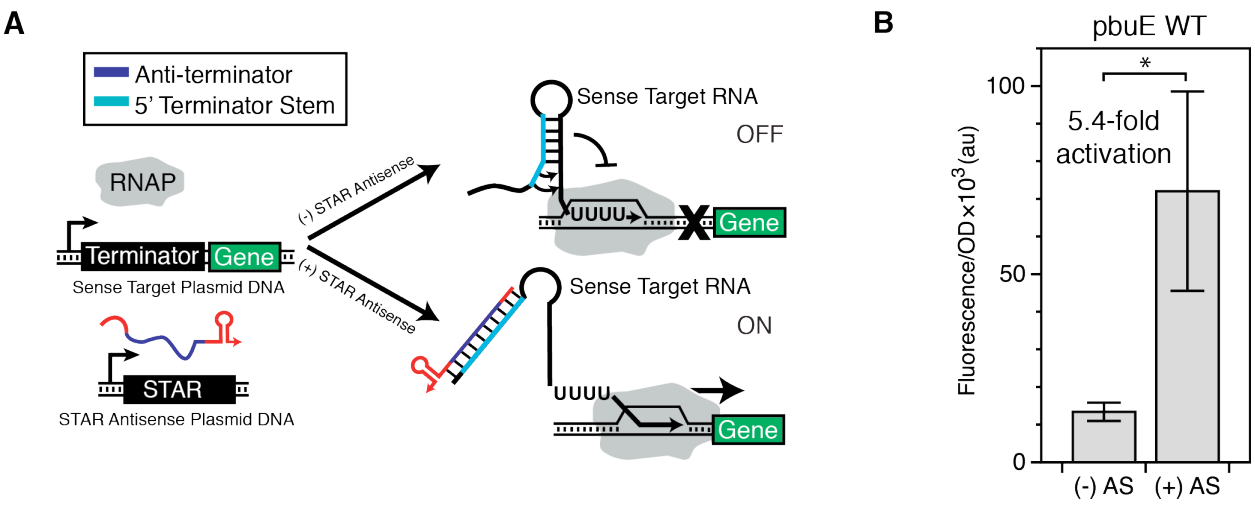


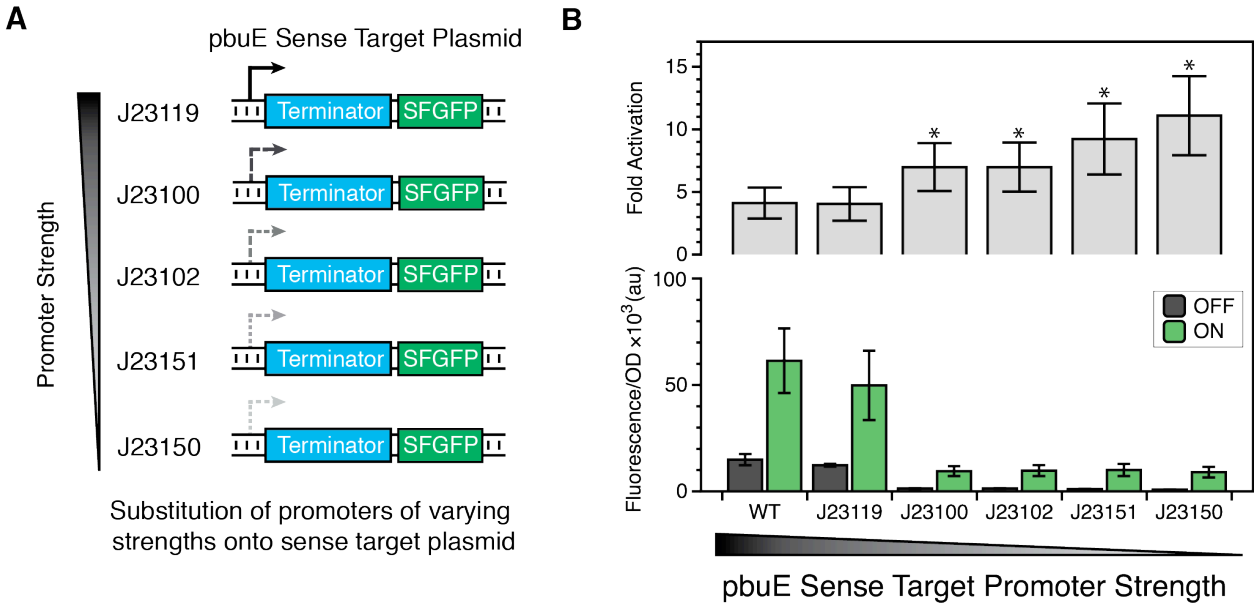
Figure 3.1 Design and function of a model Small Transcription Activating RNA (STAR). **(A)** Schematic of the mechanism following Chappell et al. (2015). In the absence of the STAR antisense, the nascent sense target RNA upstream of the reporter gene forms a terminator hairpin that stops transcription by RNA polymerase (RNAP) before the gene is transcribed (OFF). The STAR antisense is designed to contain an anti-terminator sequence complementary to the 5' side of the terminator hairpin. When the STAR antisense is present, it binds to the terminator sequence, preventing formation of the terminator hairpin and allowing transcription of the downstream gene (ON). **(B)** *In vivo* characterization of the pbuE STAR regulator performed in this study using superfolder GFP (SFGFP) fluorescence to measure gene expression from a sense target plasmid with and without a STAR antisense (AS) plasmid. Normalized fluorescence was divided by OD_{600} (Optical density at 600 nm) to give Fluorescence/OD, and fold activation was calculated as Fluorescence/OD ON divided by Fluorescence/OD OFF. Error bars represent sample standard deviation over 3 independent replicates with 3 colonies each (n=9). The * symbol indicates a statistically significant ($p<0.05$) increase in Fluorescence/OD in the case with STAR antisense as determined by a two-sided T-test.

3.4.2 Improving fold activation of STARs by manipulating STAR/target expression ratios

To begin, we chose to investigate improving fold activation by increasing the relative concentration ratio of STAR antisense to its complementary sense target. A previously developed model of the STAR mechanism hypothesized that transcription activation is directly related to the rate of binding between STAR and target (Chappell et al., 2015). Therefore, increasing the expression level of the STAR antisense relative to its target should naturally increase the number of binding events, and thus increase the likelihood of any given target RNA being transcriptionally activated. To test this, we manipulated the STAR/target expression ratio *in vivo* in *E. coli* cells by altering the relative strengths of the constitutive promoters that drive the expression of the STAR antisense and sense target in our two-plasmid system (see Materials and Methods). Given that both the STAR and target RNAs were originally under the control of the same strong constitutive σ^{70} promoter, decreasing the strength of the sense target RNA promoter provided a straightforward way to titrate down the steady-state levels of sense target RNA by reducing the transcription rate. Following this strategy, we cloned a series of successively weaker constitutive promoters upstream of the pbuE sense target RNA and examined their effects on fold activation *in vivo* (Figure 3.2). We chose promoters from the Anderson promoter library from the Registry of Standard Biological Parts (partsregistry.org), whose strengths have been well-characterized in previous work (Kelly et al., 2009) (Table B.2). The *in vivo* testing of these target promoter variants indicated that weakening the sense plasmid promoter strength did indeed result in greater fold activation, and we observed a clear correlation between decreased sense promoter strength and increased fold activation (Figure 3.2B),

with the majority of changes in fold activation being statistically significant (Table B.3). Although the overall ON level of fluorescence decreased as expected with a weaker promoter on the sense target reporter plasmid, we observed an even greater decrease in the OFF level that followed the same relative order as promoter strength (Kelly et al., 2009). This greater decrease in OFF level led to an overall increase in fold activation. When compared to the original promoter configuration, the weakest promoter tested, J23150, more than doubled the fold activation of the pbuE STAR regulator from 4.1-fold (±1.2) to 11.1-fold (±3.2). To further confirm that fold activation was directly related to the relative ratio of STAR to the sense target RNA, we characterized fold activation from the original pbuE sense target RNA as a function of decreased expression of the STAR using the same promoter series (Figure B.2). As expected, lowering the STAR expression resulted in lower fold activation, confirming the importance of a high relative ratio of STAR antisense RNA to sense target RNA for high fold activation. Overall, our results demonstrate that we can increase the fold activation of STAR regulators through manipulating STAR/target expression ratios.

Figure 3.2 Optimization of the STAR/target expression ratio yields higher fold activation. **(A)** Schematic of the design strategy for improved activation. A number of weaker promoters were substituted for the strong wild-type (WT) promoter (J23119 SpeI) on the sense target plasmid in order to decrease the expression level of the sense RNA. In these experiments, the STAR was expressed from a high-copy plasmid using the strong J23119 SpeI promoter. This promoter series was designed to increase the relative expression ratio of STAR to target. **(B)** *In vivo* fluorescence characterization demonstrates an increase in fold activation when a weaker target promoter is used, with a fold activation of 11.1-fold (±3.6) observed for the weakest J23150 promoter. Normalized fluorescence divided by optical density at 600 nm (Fluorescence/OD) is plotted on the lower axes, with the dark gray bars representing the OFF level (no-STAR control plasmid) and the green bars representing the ON level (STAR present). Fold activation (Fluorescence/OD ON divided by Fluorescence/OD OFF) is plotted on the upper axes as a series of light gray bars. Error bars represent sample standard deviation over 3 independent replicates with 3 colonies each (n=9). The * symbol indicates statistically significant ($p<0.05$) improvement in fold activation compared to the wild-type sense promoter configuration (J23119 SpeI) as determined by a two-sided T-test. Statistical significances of fold activation changes between different promoters are presented in Table B.3.

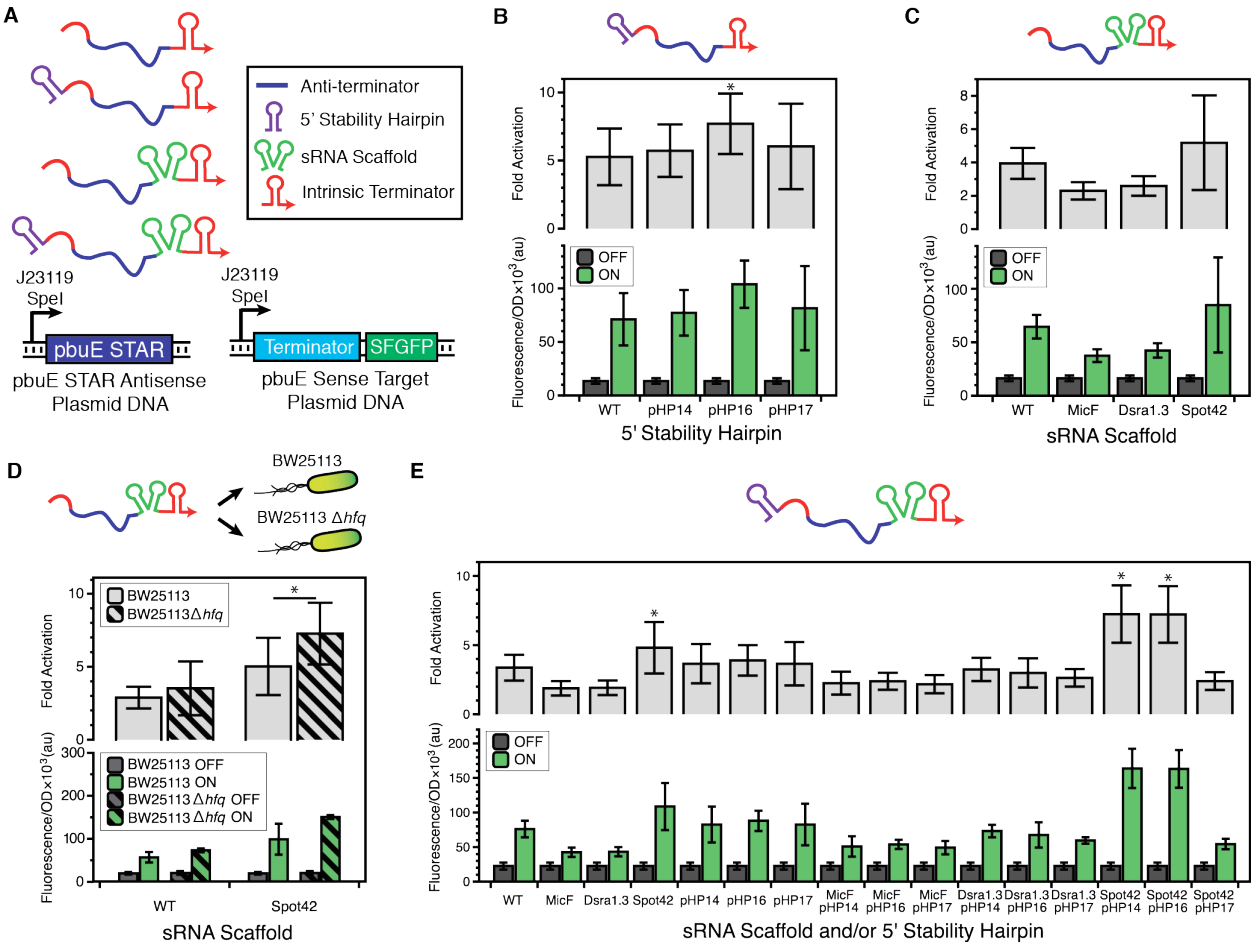

A
pbuE Sense Target Plasmid
Promoter Strength
J23119
J23100
J23102
J23151
J23150
Terminator
SFGFP
Substitution of promoters of varying strengths onto sense target plasmid
B
Fold Activation
15
10
5
0
Fluorescence/OD ×10^3 (au)
100
50
0
OFF
ON
WT
J23119
J23100
J23102
J23151
J23150
pbuE Sense Target Promoter Strength

3.4.3 Improving fold activation of STARs by engineering the STAR antisense

Since manipulating RNA stability is a key point of control for RNA mechanisms (Chappell et al., 2013; Smolke and Keasling, 2002), we next sought to examine how altering the stability of the STAR antisense could be used as another strategy for improving fold activation. The steady-state level of STAR antisense molecules available for activation at any time is governed by the balance between two key rates: the rate of synthesis (transcription) and the rate of degradation. Since the STAR antisense was already expressed from a high-copy plasmid (ColE1 origin of replication) under the control of a strong promoter, reducing the degradation rate presented a more accessible way to increase the level of STAR antisense and thus fold activation. We sought to decrease the degradation rate of the STAR antisense by adding strong RNA hairpins to the 5' end, following previous work using these hairpins to stabilize mRNAs (Carrier and Keasling, 1999) (Figure 3.3A).

Our STAR stabilization strategy was based on the fact that secondary structures located at the 5' end of bacterial mRNAs can confer stability (Carrier and Keasling, 1997; Emory et al., 1992) by blocking RNAse E-mediated degradation. In particular, this strategy has been shown to improve mRNA stability and lengthen mRNA half-lives (Carrier and Keasling, 1999). To investigate whether or not these stability hairpins added to STARs would improve fold activation, we added three previously published synthetic RNA stability hairpins (pHP14, pHP16, and pHP17) (Carrier and Keasling, 1999) to the 5' end of the pbuE STAR. We then characterized the resulting fold activation of these modified STARs (Figure 3.3B). Compared to the initial wild-type (WT) pbuE STAR, the variant with the added pHP16 5' stability hairpin demonstrated modest, yet statistically significant,

Figure 3.3 Improving fold activation by engineering the STAR antisense. **(A)** Schematic of the changes made to the pbuE STAR. Stability hairpins (purple) were added to the 5' end of the pbuE STAR to block RNase E-mediated degradation (Carrier and Keasling, 1997). sRNA scaffolds (green) were added to the 3' end to improve sRNA function, as previously demonstrated by Sakai et al. (2013). The J23119 SpeI promoter was used to express the pbuE STAR variant and sense target RNAs. Data is plotted as in Figure 3.2. **(B)** *In vivo* functional characterization of stability hairpin STAR variants indicates that the addition of the pHP16 5' stability hairpin confers a modest increase in activation function to the pbuE STAR regulator. **(C)** *In vivo* functional characterization of sRNA scaffold pbuE STAR variants indicates that only the addition of the Spot42 sRNA scaffold results in a small increase in fold activation, though this is not statistically significant over the initial wild-type (WT) pbuE STAR. Addition of the MicF or Dsra1.3 sRNA scaffolds to the pbuE STAR was shown to decrease fold activation. **(D)** Testing of the Spot42 sRNA scaffold variant of the pbuE STAR in BW25113 and BW25113 Δ*hfq* strains demonstrates slightly *higher* activation levels in the absence of Hfq. **(E)** Combining both sRNA scaffolds and 5' stability hairpins yields the highest fold activation for the combinations of Spot42/pHP14 and Spot42/pHP16. In parts **(B)**, **(C)**, and **(E)** the * symbol indicates a statistically significant ($p<0.05$) increase in fold activation over WT as determined by a two-sided T-test. In part **(D)** * indicates a statistically significant ($p<0.05$) difference in fold activation between the two strains tested as determined by a two-sided T-test.


A
Anti-terminator
5' Stability Hairpin
sRNA Scaffold
Intrinsic Terminator
J23119
SpeI
pbuE STAR
pbuE STAR Antisense
Plasmid DNA
J23119
SpeI
Terminator
SFGFP
pbuE Sense Target
Plasmid DNA
B
Fold Activation
Fluorescence/OD×10³(au)
OFF
ON
WT
pHP14
pHP16
pHP17
5' Stability Hairpin
C
Fold Activation
Fluorescence/OD×10³(au)
OFF
ON
WT
MicF
Dsra1.3
Spot42
sRNA Scaffold
D
BW25113
BW25113 Δhfq
Fold Activation
BW25113
BW25113Δhfq
Fluorescence/OD×10³(au)
BW25113 OFF
BW25113 ON
BW25113 Δhfq OFF
BW25113 Δhfq ON
WT
Spot42
sRNA Scaffold
E
Fold Activation
Fluorescence/OD×10³(au)
OFF
ON
WT
MicF
Dsra1.3
Spot42
pHP14
pHP16
pHP17
MicF pHP14
MicF pHP16
MicF pHP17
Dsra1.3 pHP14
Dsra1.3 pHP16
Dsra1.3 pHP17
Spot42 pHP14
Spot42 pHP16
Spot42 pHP17
sRNA Scaffold and/or 5' Stability Hairpin

improvements in *in vivo* fold activation, going from 5.2-fold (±2.1) for the WT pbuE STAR to 7.7-fold (±2.2) with pHP16 (Figure 3.3B). Interestingly, the level of increased activation did not directly correlate with the previously reported half-lives of these stability hairpins (Carrier and Keasling, 1999). While pHP14, pHP16, and pHP17 had been shown by Carrier and Keasling (1999) to confer successively longer half-lives (in numerical order), in the context of the pbuE STAR, only pHP16 appeared to confer a statistically significant increase in fold activation.

To test whether these hairpins conferred stability in the STAR context, we used RT-qPCR (Reverse transcription quantitative polymerase chain reaction) to quantify the relative degradation of the pHP14, pHP16 and WT pbuE STAR variants at different time points after the addition of rifampicin to the media to halt transcription (see Appendix B.3: Supplemental Methods). We found that both pHP14 and pHP16 significantly reduced degradation of pbuE compared to the WT pbuE STAR (Figure B.3) indicating that both hairpins do in fact stabilize the pbuE STAR. To understand why only modest levels of fold activation increase were observed with these hairpins, we used computational RNA structure modeling with RNAStructure (Reuter and Mathews, 2010) to predict the structures of the pbuE STAR variant with the added stability hairpins (Figure B.4). This analysis showed that the stability hairpins could alter the predicted structure of the pbuE STAR RNAs. This folding could be enough to offset the benefits of increasing STAR stability, though in the case of pHP16 the benefits of stability outweighed the possible alteration of STAR structure.

As an alternate method of increasing fold-activation through engineering the STAR antisense, we tested adding sRNA-derived scaffolds to the 3’ end of the STAR, upstream of

the transcriptional terminator. This strategy was based upon recent research (Na et al., 2013; Sakai et al., 2013; Sharma et al., 2012) showing that the addition of sRNA scaffolds to transcriptional and translational synthetic sRNA regulators can lead to improved function. Not only are the scaffolds theorized to stabilize the RNA through the addition of secondary structure, these particular sRNA-derived scaffolds are also designed to include binding sites for the RNA-binding chaperone protein Hfq. Among its many roles, Hfq is known to aid sRNA function by mediating sRNA interactions with target mRNAs for many trans-encoded sRNAs that regulate translation (Møller et al., 2002; Zhang et al., 2003). Hfq can also modulate sRNA stability by affecting ribonuclease accessibility or susceptibility to 3' polyadenylation and subsequent degradation (Vogel and Luisi, 2011). To test the ability of sRNA scaffolds to increase fold activation for STAR regulators, we chose the MicF, Dsra1.3 and Spot42 scaffolds that were previously shown to lead to the greatest improvements in activation and repression levels in the RNA regulators tested(Sakai et al., 2013). We then fused each scaffold to the 3' end of the pbuE STAR, directly before the transcriptional terminator (Figure 3.3A). Functional testing *in vivo* showed variable results depending on the scaffold used. In particular, while the addition of the Spot42 scaffold to the pbuE STAR antisense slightly increased activation, both MicF and Dsra1.3 markedly decreased activation levels (Figure 3.3C). Structural prediction with RNAStructure (Reuter and Mathews, 2010) indicated that a number of the most probable low-energy structures for the antisense fused with MicF or Dsra1.3 interfered with the native structure of both the STAR antisense and the sRNA scaffolds (Figure B.5). Thus the observed decrease in activation in the case of the MicF and Dsra1.3 scaffolds could be the result of structural interference between the scaffold and the STAR.

While the Spot42 variant fold activation was statistically different from the other two scaffold variants, it was not statistically different from the WT pbuE STAR fold activation. Despite this, we did confirm that the Spot42 sRNA scaffold increased STAR stability compared to the WT pbuE STAR, though not as much as the pHP14 and pHP16 stability hairpins (Figure B.3). Since it was the only functioning scaffold variant, we next sought to test the role of Hfq in the observed modest increase in STAR activation with the Spot42 scaffold. To test this, we repeated the *in vivo* functional characterization of the Spot42 fusions in both the Keio collection Δ*hfq* knockout strain and its parent *E. coli* K-12 BW25113 strain (Baba et al., 2006) (Figure 3.3D). We found that the absence of Hfq had little to no effect on the wild-type pbuE STAR activation, as expected given that the wild-type does not contain an Hfq-recruiting scaffold sequence. Surprisingly, we found that the activation level of the pbuE STAR-Spot42 fusion significantly improved in the absence of Hfq, in contrast to previous observations of the reliance of sRNA scaffolds on Hfq for added stability (Sakai et al., 2013).

Next, we examined whether we could increase fold activation further by combining the stability hairpin and sRNA scaffold strategies together. To test this, we created all possible combinations of the pbuE STAR antisense containing both 5' stability hairpins and 3' sRNA scaffolds. *In vivo* characterization indicated that the combination of the Spot42 scaffold with either the pHP14 or pHP16 stability hairpin granted significantly improved activation over any of the variants with only hairpin or scaffold (Figure 3.3E). In particular, we observed 7.2-fold (±2.1) activation for the variant with both Spot42 and pHP14 and 7.2-fold (±2.0) activation for the variant with Spot42 and pHP16, both significantly better than the 4.8-fold (±1.9) activation seen for Spot42 alone, though statistically indistinguishable

from each other. The addition of both a 5' stability hairpin and the Spot42 scaffold may alter the STAR structure so that it can fully benefit from the stabilization of the stability hairpin to increase overall fold activation.

Interestingly, these increases in fold activation are larger than the improvements in transcription repression seen when scaffolds were applied to the pT181 RNA-based transcriptional repressor (Sakai et al., 2013). However, Sakai et al. were able to successfully use these strategies to improve the fold activation of an RNA translational activator. These results could suggest a more general principle for antisense RNA engineering strategies being more effective for gene expression activation or could indicate these strategies are more effective when applied to relatively unstructured antisense RNAs.

Overall, our results showed that the incremental improvements in fold activation generated by the addition of 5' stability hairpins and 3' scaffolds alone can be combined to generate STARs with even higher fold activation. However, more work is needed to uncover the structural basis of non-functioning variants, as well as the synergistic effect among the functional variants in order to make this strategy more predictable.

3.4.4 A general method for increasing STAR activation by combining expression level tuning and RNA engineering strategies

We subsequently investigated whether combining expression level tuning and RNA engineering strategies could be used to further optimize fold activation for the pbuE STAR. To do this, we combined the Spot42/pHP14 pbuE antisense variant together with the weakened promoter strength series on the sense target RNA. Combining the stabilized STARs with the weaker promoter sense target plasmids yielded another increase in

activation level, reaching as high 13.4-fold (±3.8) activation in the case of the Spot42 pHP14 antisense combined with the sense target plasmid containing the J23151 promoter (Figure 3.4A). While the general trend was toward higher fold activation with weaker promoters, there was high error, and the trend was less uniform than seen previously, though the differences between the J23151 variant and all others were statistically significant. Nevertheless, the higher fold activation seen from combining the optimized pbuE STAR with weaker sense target promoters indicates that RNA engineering strategies can be combined with expression tuning to increase fold activation.

Having successfully optimized the pbuE transcriptional activator from its initial 5.3-fold (±2.2) activation level to 13.4-fold (±3.8) activation, we next sought to test the generality of this combined method for improving STAR fold activation. We started with a previously constructed STAR generated from the prgX conjugation control system (Weaver, 2007) that only displayed a 2.1-fold (±0.4) activation (Chappell et al., 2015) (Figure 3.4B). In particular, we applied the same 5' stability hairpins as used above and the Spot42 sRNA scaffold to the prgX STAR. *In vivo* testing revealed that while some of the hairpin additions modestly improved fold activation (Figure 3.4B), the addition of the Spot42 scaffold did not (Figure B.6). When combined with weaker sense target promoters, the best prgX STAR variant (pHP14) had an even greater increase in fold activation (Figure 3.4C), providing further proof of the modularity of the strategies for modifying sense promoter and STAR antisense stability. Overall, the best prgX variant displayed 14.6-fold (±3.7) activation, a vast improvement from the original 2.1-fold (±0.5) activation and a validation that the optimization strategies work on an additional, unrelated STAR system.

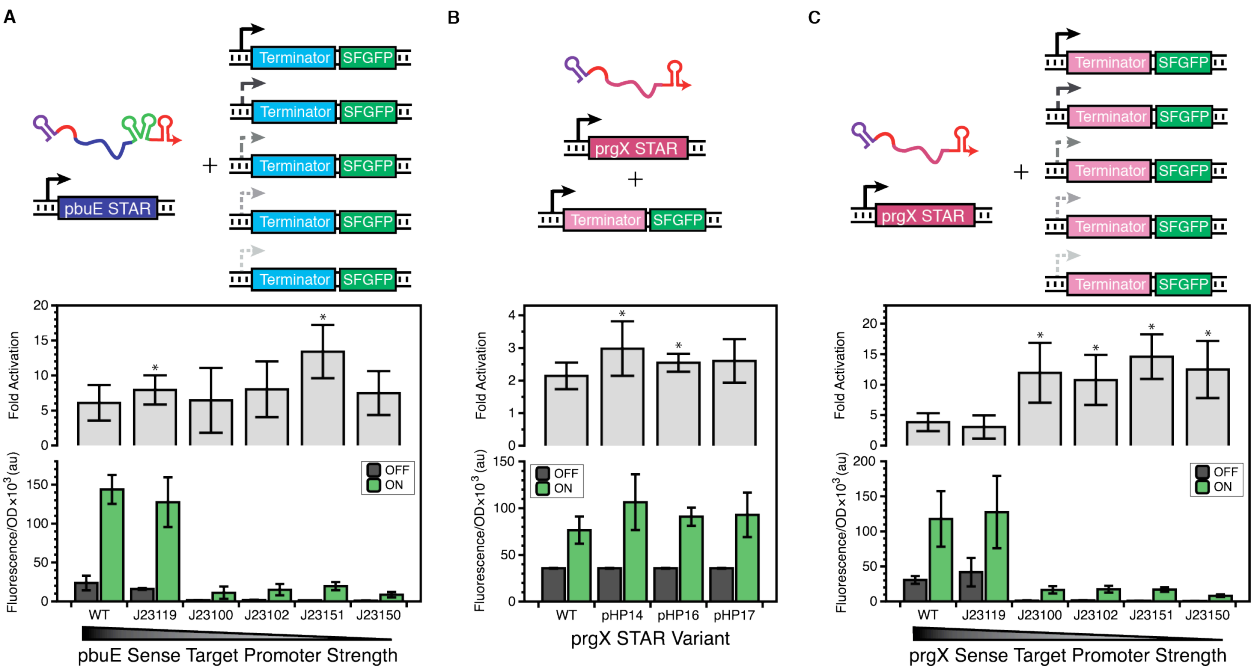


Figure 3.4 Combining expression level tuning and RNA engineering strategies improves STAR fold activation in multiple systems. Data is plotted as in Figure 3.2. **(A)** Combining the pbuE STAR antisense fused with the Spot42 scaffold and the pHP16 stability hairpin with weaker target sense promoters yields up to 13.4-fold (±2.0) activation for the J23151 promoter pairing. **(B)** 5' stability hairpins slightly improve fold activation for the prgX STAR regulator. **(C)** Combining the prgX STAR fused with the pHP14 stability hairpin with weaker strength target promoters yields another boost in fold activation, showing the broader applicability of these STAR optimization strategies. Wild-type (WT) indicates the initial pbuE STAR regulator in part **(A)** and the initial prgX STAR regulator in part **(B)** and **(C)**, both of which use the strong J23119 SpeI promoter to drive both STAR and target RNAs. The * indicates a statistically significant ($p<0.05$) increase in fold activation over WT, as determined by a two-sided T-test.

3.4.5 Testing the orthogonality of optimized STARs to each other and to RNA transcriptional repressors

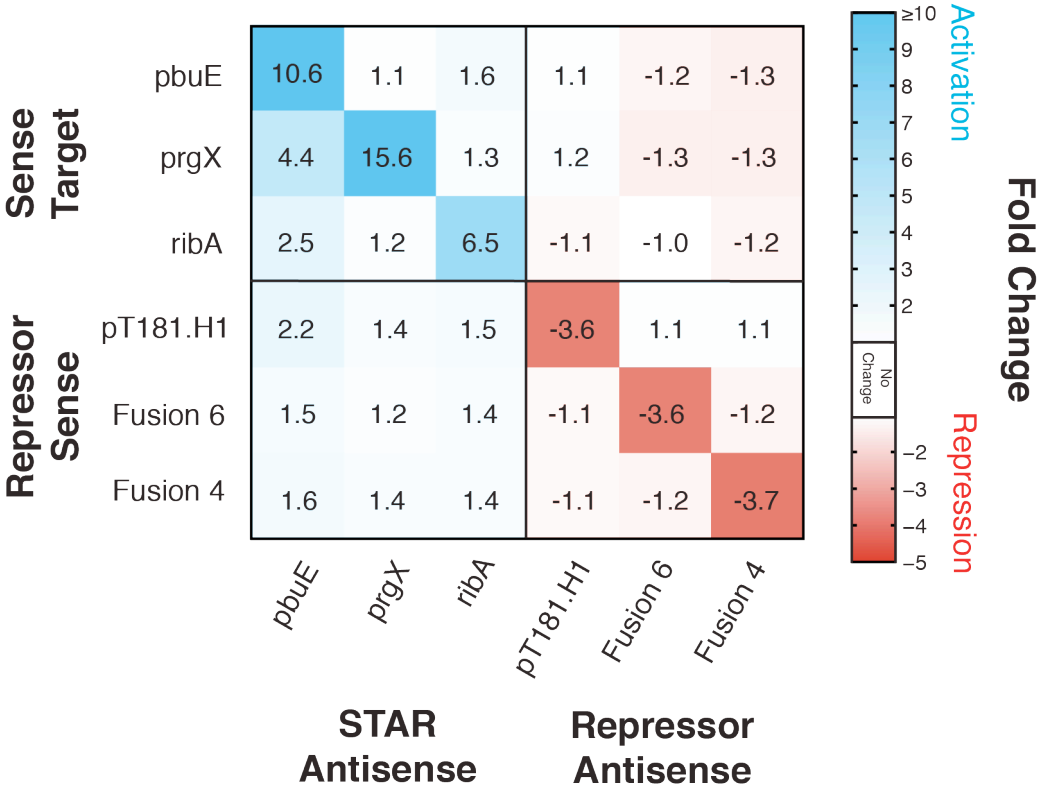
We next sought to test whether the best variants of our two newly optimized STARs were orthogonal to the previously reported ribA STAR (Chappell et al., 2015) and three previously developed transcriptional repressors (Lucks et al., 2011; Takahashi and Lucks, 2013). If two regulators are orthogonal, then the antisense of one should not regulate the sense target RNA of the other (and vice versa), allowing them to be used together in a complex regulatory system without cross-talk. Such orthogonality is non-trivial, especially given that the inclusion of additional RNA sequence within the optimized STARs increases the potential for off-target interactions, as does increasing the concentration ratios of STARs to their targets. Moreover, orthogonality of these component parts is key for their use in the higher-order logic gates and circuits that STARs and transcriptional repressors have been used to construct (Chappell et al., 2015; Lucks et al., 2011; Takahashi and Lucks, 2013). For example, the ability to build circuits using orthogonal elements allows synthetic biologists to program systems with complex functionalities like ligand-sensitive NOR gates (Qi et al., 2012b) and RNA cascades that control the timing of gene expression (Takahashi et al., 2014), a vital capability for biotechnology applications.

To perform the orthogonality test, we challenged each optimized STAR antisense variant against the sense targets from the pbuE and prgX systems (using the J23151 promoter), along with the sense target from the previously reported ribA STAR (Chappell et al., 2013). We also checked for orthogonality to sense target regions from three RNA transcriptional repressors: pT181.H1 from Lucks et al. (2011), Fusion 4, and Fusion 6 from

Takahashi and Lucks (2013). As a further test, we challenged each RNA transcriptional repressor antisense variant against the same set of sense targets. This resulted in a 6x6 matrix of conditions demonstrating the observed orthogonality of these regulators to one another (Figure 3.5, Figure B.7). Both of the newly optimized STARs showed reasonable orthogonality to the other regulators. We did observe that the only off-target activation was between the pbuE STAR antisense and the prgX, ribA and pT181.H1 sense targets, though this effect was well below the activation seen with its cognate target, and was associated with high experimental variability (Figure B.7). More work is needed to understand the nature of the transient interactions that may occur between STARs and their potential targets during the dynamic process of transcription to understand this effect.

These results confirm that these STAR optimization strategies largely do not affect the orthogonality of STARs between themselves and RNA transcriptional repressors. Not only will the addition of new orthogonal STARs allow for more complex RNA circuitry, the optimization strategies used to improve STAR activation will allow for the future development of more highly functional RNA transcriptional activators.

Figure 3.5 Testing the orthogonality of improved STAR regulators with each other and with RNA transcriptional repressors. Characterization of a 6 x 6 orthogonality matrix consisting of the newly optimized pbuE and prgX STAR regulators, the previously reported ribA STAR regulator (Chappell et al., 2015), the pT181.H1 repressor (Lucks et al., 2011) and Fusion 4 and Fusion 6 transcriptional repressors reported by Takahashi and Lucks (2013). Each matrix square represents the fold change of gene expression for the indicated combination of STAR or repressor plasmid and target plasmid, compared to the condition with target plasmid and an empty no-antisense control plasmid. Fluorescence characterization (measured in units of Fluorescence/OD, fluorescence divided by optical density at 600 nm) was used to calculate average fold change, which is represented by a color scale in which ≥ 10-fold is the darkest blue (activation), 1-fold is white (no activation or repression) and -5-fold is red (repression). Fluorescence/OD plots for each individual combination are shown in Figure B.7. Data represents mean values of n = 9 biological replicates.


Sense Target
Repressor Sense
pbuE
prgX
ribA
pT181.H1
Fusion 6
Fusion 4
10.6
1.1
1.6
1.1
-1.2
-1.3
4.4
15.6
1.3
1.2
-1.3
-1.3
2.5
1.2
6.5
-1.1
-1.0
-1.2
2.2
1.4
1.5
-3.6
1.1
1.1
1.5
1.2
1.4
-1.1
-3.6
-1.2
1.6
1.4
1.4
-1.1
-1.2
-3.7
pbuE
prgX
ribA
pT181.H1
Fusion 6
Fusion 4
STAR Antisense
Repressor Antisense
Fold Change
Activation
Repression
No Change
≥10
9
8
7
6
5
4
3
2
−2
−3
−4
−5

3.5 Conclusions

In this work, we tested several RNA engineering strategies for optimizing the fold activation of small transcription activating RNAs. In particular, we focused on strategies designed to stabilize and improve the STAR antisense design and alter the concentration ratio between the STAR and its target. Using the pbuE STAR as a test case, we showed that the addition of 5' stability hairpins and sRNA scaffolds to the STAR antisense and the ability to adjust the ratio of STAR antisense to target sense via promoter strength tuning gave convenient and sometimes modular ways to alter the transcription activation levels of two distinct STAR systems. Specifically, we found that these strategies as applied to the pbuE STAR system increased the fold activation from 5.3-fold (±2.2) to 13.4-fold (±3.8). Moreover, we showed that these strategies were general and, when applied to the unrelated prgX STAR regulator, yielded an increase in transcription activation from 2.1-fold (±0.4) to 14.6-fold (±3.7). Furthermore, we showed that these changes largely preserved the orthogonality of the optimized STARs to each other and to a panel of RNA transcriptional repressors that have been used to construct higher-order RNA transcriptional circuits (Chappell et al., 2015; Lucks et al., 2011).

These results are significant for several reasons. First, these optimizations have expanded the repertoire of STARs - the starting point fold activations for the pbuE and prgX STAR regulators prohibited their use for higher order circuit construction, a problem remedied by our optimizations. Second, the optimization strategies used on the pbuE and prgX systems should be applicable to many other STARs, paving the way for even larger libraries of RNA regulators. Finally, the demonstrated orthogonality between the newly

optimized activators and previously reported regulators, in addition to representing a non-trivial achievement, makes them highly useful for future circuit-building.

It should be noted that the majority of the improvement in STAR fold activation was achieved through decreasing the OFF level. If a high ON level were required for a particular application, other strategies could be used to increase ON levels while maintaining low OFF levels. For instance, if high protein expression were desired, modular strategies aimed at tuning translation through altering ribosome binding site strength and accessibility could be used as an alternate way of manipulating target gene expression levels (Salis et al., 2009).

In addition to optimizing STAR fold activation, the tested strategies have led to a series of STARs with varying ON, OFF and fold activation levels. These strategies have thus created a panel of variants that can be used for fine-tuning of transcription activation. The importance of fine-tuning individual regulators to enable the correct performance of a larger circuit has been demonstrated in numerous examples of synthetic circuits (Ellis et al., 2009; Elowitz and Leibler, 2000; Wang et al., 2009), making the suite of functional STAR variants a useful library to draw from for future circuit design. Strategies like these are becoming more important as synthetic biology looks to implement increasingly sophisticated genetic circuitry, requiring the ability to carefully tune the biological circuit components.

In summary, we have successfully expanded our capabilities for genetic regulation and made highly useful additions to the synthetic biology toolkit through systematic optimizations of a set of small transcription activating RNAs. These newly-improved STAR regulators will allow for the construction of complex cellular circuitry, and the optimization

strategies will be highly useful for creating a generation of additional STARs for use in a broad range of biotechnologies.

3.6 Acknowledgements

We thank Professor Matthew DeLisa (Cornell Chemical and Biomolecular Engineering) for providing the BW25113 and BW25113 Δhfq strains. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program [DGE-1144153 to S.M.], Defense Advanced Research Projects Agency Young Faculty Award (DARPA YFA) [N66001-12-1-4254 to J.B.L.], an Office of Naval Research Young Investigators Program Award (ONR YIP) [N00014-13-1-0531 to J. B. L.], and an NSF CAREER award [1452441 to J. B. L.] J.B.L. is an Alfred P. Sloan Research Fellow.

The authors declare competing financial interest. The authors have submitted a provisional patent application (No. 61/981,241) for the technologically important developments included in this article.

3.7 Bibliography

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 1–11.

Bhadra, S., and Ellington, A.D. (2014). Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. Nucleic Acids Res. *42*, e58.

Brantl, S., and Wagner, E.G. (2000). Antisense RNA-mediated transcriptional attenuation:

an in vitro study of plasmid pT181. Mol. Microbiol. *35*, 1469–1482.

Carrier, T.A., and Keasling, J.D. (1997). Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. Biotechnol. Prog. *13*, 699–708.

Carrier, T.A., and Keasling, J.D. (1999). Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. Biotechnol. Prog. *15*, 58–64.

Cech, T.R., and Steitz, J.A. (2014). The Noncoding RNA Revolution — Trashing Old Rules to Forge New Ones. Cell *157*, 77–94.

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013a). Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth. Biol. *2*, 463–472.

Ceres, P., Trausch, J.J., and Batey, R.T. (2013b). Engineering modular "ON" RNA switches using biological components. Nucleic Acids Res. *41*, 10449–10461.

Chappell, J., Takahashi, M.K., Meyer, S., Loughrey, D., Watters, K.E., and Lucks, J. (2013). The centrality of RNA for engineering gene expression. Biotechnol. J. *8*, 1379–1395.

Chappell, J., Takahashi, M.K., and Lucks, J.B. (2015). Creating small transcription activating RNAs. Nat. Chem. Biol. *11*, 1–9.

Collins, J.A., Irnov, I., Baker, S., and Winkler, W.C. (2007). Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev. *21*, 3356–3368.

DebRoy, S., Gebbie, M., Ramesh, A., Goodson, J.R., Cruz, M.R., van Hoof, A., Winkler, W.C., and Garsin, D.A. (2014). A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science *345*, 937–940.

Ellis, T., Wang, X., and Collins, J.J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat. Biotechnol. *27*, 465–471.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338.

Emory, S.A., Bouvet, P., and Belasco, J.G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. Genes Dev. *6*, 135–148.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. *9*, 102–114.

Gottesman, S., and Storz, G. (2011). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. Cold Spring Harb. Perspect. Biol. *3*.

Green, A.A., Silver, P.A., Collins, J.J., and Yin, P. (2014). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. Cell *159*, 1–15.

Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. *22*, 841–847.

Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. J. Biol. Eng. *3*, 4.

Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D., and Arkin, A.P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc. Natl. Acad. Sci. *108*, 8617–8622.

Lynch, S.A., Desai, S.K., Sajja, H.K., and Gallivan, J.P. (2007). A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. Chem. Biol. *14*, 173–184.

Møller, T., Franch, T., Højrup, P., Keene, D.R., Bächinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002). Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol. Cell *9*, 23–30.

Mutalik, V.K., Qi, L., Guimaraes, J.C., Lucks, J.B., and Arkin, A.P. (2012). Rationally designed families of orthogonal RNA regulators of translation. Nat. Chem. Biol. *8*, 447–454.

Na, D., Yoo, S.M., Chung, H., Park, H., Park, J.H., and Lee, S.Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. Nat. Biotechnol. *31*, 170–174.

Nou, X., and Kadner, R.J. (2000). Adenosylcobalamin inhibits ribosome binding to btuB RNA. Proc. Natl. Acad. Sci. *97*, 7190–7195.

Pardee, K., Green, A.A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., and Collins, J.J. (2014). Paper-Based Synthetic Gene Networks. Cell *159*, 940–954.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K., and Arkin, A.P. (2012a). Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res. 1–12.

Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K., and Arkin, A.P. (2012b). Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res. *40*, 5775–5786.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rodrigo, G., Landrain, T.E., and Jaramillo, A. (2012). De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. Proc. Natl. Acad. Sci. *109*, 15271–15276.

Sakai, Y., Abe, K., Nakashima, S., Yoshida, W., Ferri, S., Sode, K., and Ikebukuro, K. (2013). Improving the Gene-Regulation Ability of Small RNAs by Scaffold Engineering in Escherichia coli. ACS Synth. Biol. *3*, 152–162.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. *27*, 946–950.

Sharma, V., Yamamura, A., and Yokobayashi, Y. (2012). Engineering artificial small RNAs for Conditional gene silencing in escherichia coli. ACS Synth. Biol. *1*, 6–13.

Smolke, C.D., and Keasling, J.D. (2002). Effect of copy number and mRNA processing and stabilization on transcript and protein levels from an engineered dual-gene operon. Biotechnol. Bioeng. *78*, 412–424.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol. Cell *43*, 880–891.

Takahashi, M.K., and Lucks, J.B. (2013). A modular strategy for engineering orthogonal chimeric RNA transcription regulators. Nucleic Acids Res. *41*, 7577–7588.

Takahashi, M.K., Chappell, J., Hayes, C.A., Sun, Z.Z., Kim, J., Singhal, V., Spring, K.J., Al-Khabouri, S., Fall, C.P., Noireaux, V., et al. (2014). Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. ACS

Synth. Biol.

Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. *9*, 578–589.

Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F., and Mörl, M. (2013). De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res. *41*, 2541–2551.

Wang, H.H., Isaacs, F.J., Carr, P. a, Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.

Weaver, K.E. (2007). Emerging plasmid-encoded antisense RNA regulated systems. Curr. Opin. Microbiol. *10*, 110–116.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-Input RNAi-Based Logic Circuit. Science. *333*, 1307–1311.

Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of Intrinsic Transcription Termination and Antitermination. Science *284*, 611–615.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. *50*, 1111–1124.

CHAPTER 4

ENGINEERING LIGAND-RESPONSIVE RNA ACTIVATORS

4.1 Abstract

Generating biological circuits capable of sensing their environment is one of the key goals of synthetic biology. With this in mind, I sought to expand upon our previous work developing small transcription activating RNAs (STARs) (see Chapter 3) by making them responsive to externally supplied ligands through fusion to an RNA aptamer domain. We engineered these candidate "aptaSTARs" through a variety of strategies, including randomized screening and computational design, eventually isolating a handful of ligand-sensitive variants that demonstrate 2 to 3-fold activation in response to the presence of ligand. Although many applications may require greater fold activation, these preliminary results demonstrate the feasibility of the aptaSTAR concept and re-affirm the flexibility of RNA regulation.

4.2 Introduction

Natural and artificial RNA regulators offer synthetic biologists important tools for manipulating gene expression, allowing the construction of biological circuitry with lower metabolic burden than circuits built from protein regulators. The recent advent (Chappell et al., 2015) and optimization (Meyer et al., 2016) of small transcription activating RNAs (STARs) filled a gap in the suite of naturally-occurring regulators by generating small RNAs that activate transcription *in trans*, thus further expanding the regulatory capabilities of RNA. Now we seek to take these STARs one step further and make them ligand-responsive

by fusing them to RNA aptamers, segments of RNA that bind a particular target ligand. These newly created "aptaSTAR" fusions will allow for greater complexity and responsiveness in RNA-only circuitry.

Of course, ligand-sensing RNAs in the form of natural and synthetic riboswitches are already well represented in the literature (Chappell et al., 2013; Winkler and Breaker, 2005). The wide-spread occurrence of riboswitches within bacterial genomes (Kim et al., 2015; Mandal and Breaker, 2004; Mandal et al., 2004; Serganov et al., 2006) underscores the importance and utility of linking gene expression to the surrounding environment, and the engineering of synthetic riboswitches has also been a topic of active research in the synthetic biology community (Borujeni et al., 2015; Ceres et al., 2013a; Lynch and Gallivan, 2009; Mishler and Gallivan, 2014). Work by Ceres et al. (2013a, 2013b) generated synthetic riboswitches by combining naturally occurring riboswitch expression platforms with alternate aptamer domains in a carefully-designed manner, while other research has focused on isolating functional riboswitches by screening vast libraries of potential sequences in which key linker regions were varied to allow for ligand-mediated structural transitions (Lynch and Gallivan, 2009; Lynch et al., 2007).

AptaSTARs would add to this pre-existing toolbox of small-molecule sensors and even allow for more complex circuit construction—because aptaSTARs act *in trans* and depend upon the presence of two elements (ligand and aptaSTAR) in order to activate the target gene, they allow for higher-level conditional control. AptaSTAR-based circuits could tie the transcription of the aptaSTAR itself to an independent contingency apart from the presence of ligand, allowing for the creation of AND gates and other logical operators.

In our efforts to engineer aptaSTARs, we relied on a number of different strategies—

ranging from rational and computational design to the screening of carefully designed randomized libraries—and sought to develop aptaSTARs with a number of different ligand specificities. The preliminary work presented here demonstrates the development of aptaSTARs with reasonable activation in response to the presence of external ligand, providing proof-of-concept for the construction of ligand-sensing STARs.

4.3 Results

4.3.1 Rational design of a theophylline-sensitive RNA activator

My initial attempts at aptaSTAR design used the well-characterized theophylline aptamer, which has proved amenable to engineering in the past (Qi et al., 2012; Stojanovic and Kolpashchikov, 2004; Wachsmuth et al., 2013). In addition to being highly selective,

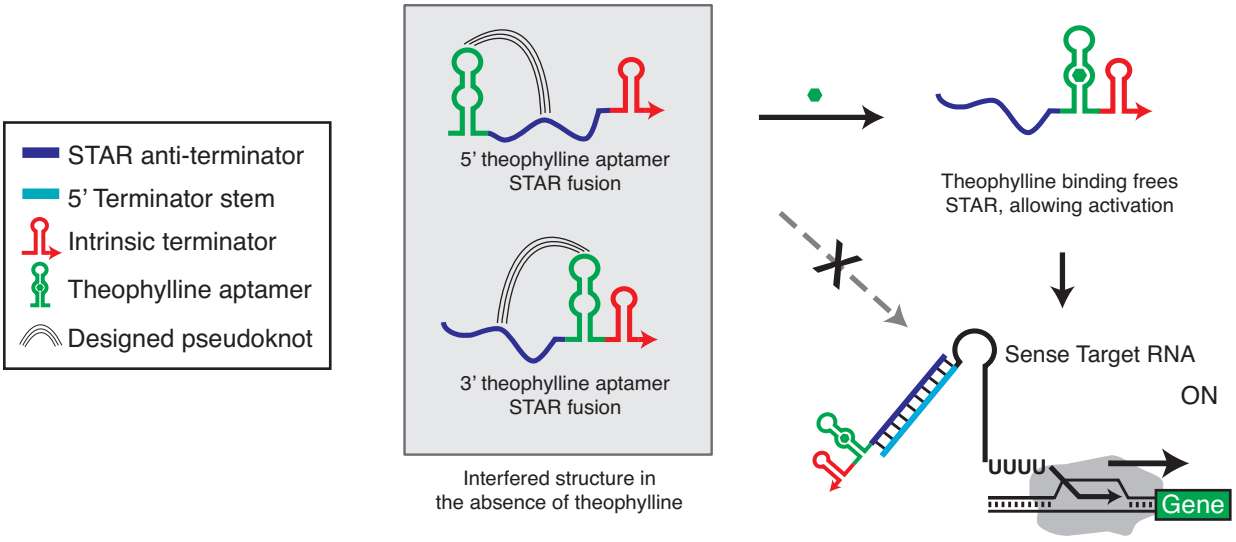


Figure 4.1 Schematic of a designed pseudoknot aptaSTAR. In the absence of theophylline, a designed pseudoknot forms between the theophylline aptamer and STAR anti-terminator, but theophylline binding disrupts the pseudoknot, freeing the STAR for trans-activation of its sense target RNA.

easily discriminating between theophylline and its close analog, caffeine, the theophylline aptamer binds theophylline through interactions with nucleotides within an inner loop region in its stem-loop structure (Figure C.1), leaving the loop region free for engineering without affecting the aptamer's affinity for theophylline (Zimmermann et al., 1997). Exploiting this loop sequence flexibility in a manner following work by Qi et al. (2012), I designed a series of potential aptaSTARs in which the theophylline aptamer was fused to either the 3' or 5' end of the STAR antisense, with a designed pseudoknot between the aptamer loop region and a 13-nucleotide section of the STAR antisense (see Figure 4.1). The library of aptaSTARs was designed so that the 13-nucleotide window of interference was walked along the length of the STAR anti-terminator, generating 24 potential designs

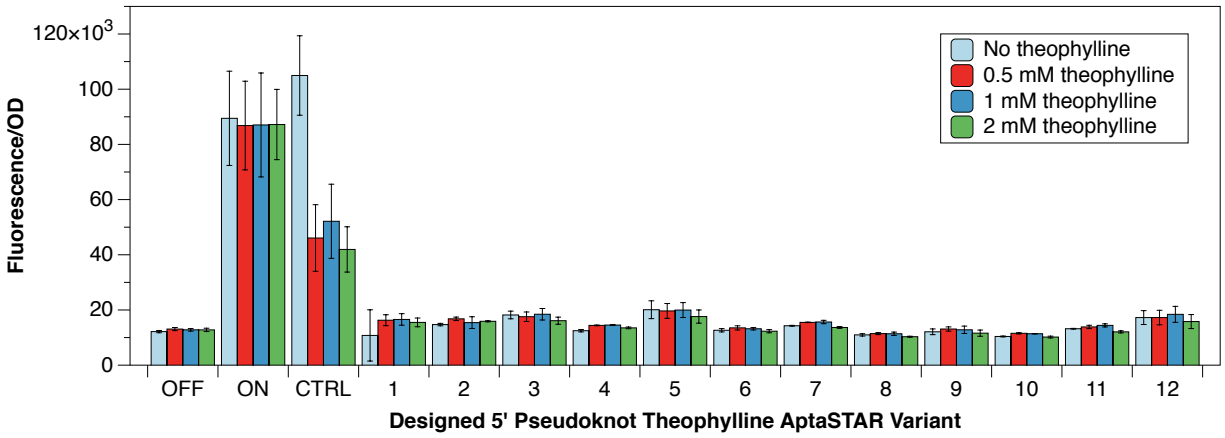


Figure 4.2 *In vivo* testing of theophylline aptaSTARs with designed 5' pseudoknot. OFF and ON show the behavior of the original pT181 STAR system without and with STAR antisense, respectively, and CTRL denotes a previously-reported theophylline-sensitive repressor. AptaSTAR variants 1-12 are designed so that the pseudoknot binding location progresses along the STAR anti-terminator region from 3' to 5'.

based on the particular pT181-based STAR used as a starting point.

The potential theophylline aptaSTARs were tested *in vivo* alongside the original pT181-based STAR and the control theophylline-sensitive repressor previously reported by Qi et al. (2012) to assay theophylline-sensitive activation. However, as shown in Figure 4.2, aptaSTARs with the designed 5' pseudoknot failed to activate even in the presence of 2 mM theophylline, and the same proved true of the other designs tested. It seems that the STAR anti-terminator is particularly sensitive to structural interference and that the pseudoknotted interaction was too strong to be easily disrupted by ligand binding, leaving the STAR in an interfered state and unable to activate transcription. Faced with the difficulty of rationally designing an interaction of the optimal strength to interfere with STAR function, yet be easily disrupted by ligand binding, I turned to functional screening as a way to test a larger potential sequence space.

4.3.2 Screening for a theophylline-sensitive RNA activator

Taking my cues from previous work demonstrating the use of high through-put screening to identify a synthetic translational riboswitch (Lynch et al., 2007), I experimented with using a randomized transducer to provide sequence flexibility for signal propagation from the theophylline aptamer to the STAR antisense. In my aptaSTAR screening libraries, I fused the theophylline aptamer to previously generated STARs (pT181, pbuE, and pAD1) via an 11-nucleotide randomized transducer region linking the theophylline aptamer to the 5' end of the STAR or to the STAR's 3' end before the intrinsic terminator (see Figure 4.3). I then subjected these randomized libraries to functional fluorescence screening, picking individual colonies from bacteria transformed with the ran-

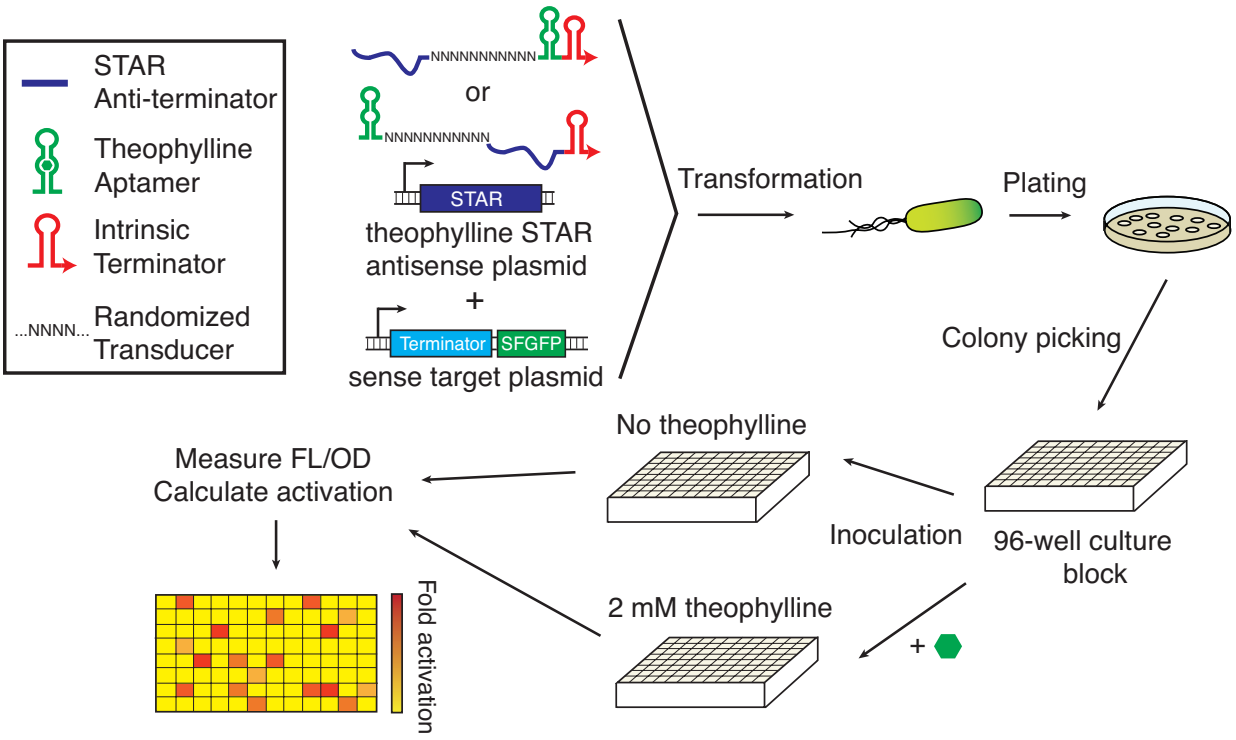


Figure 4.3 Screening for a theophylline-sensitive aptaSTAR. The theophylline aptamer was connected to a STAR via a randomized transducer region, and the resultant aptaSTAR library was then screened for functionality using a 96-well plate fluorescence assay to compare activation with and without theophylline.

domized aptaSTAR library and the sense plasmid containing SFGFP under control of the target transcriptional terminator. The colonies were then grown with and without theophylline and assayed for normalized fluorescence alongside wild-type STAR controls and a known theophylline-responsive repressor, previously reported by Qi et al. (2012).

During the screening of ~1600 colonies from the variously constructed libraries, a number of variants showed modest levels of activation in response to theophylline. The best variant, isolated from the library with the theophylline aptamer linked to the 5' end of the pbuE STAR with pHP17 stability hairpin, gave 2.7-fold activation in response to

theophylline (Figure 4.4), demonstrating that screening a randomized library can successfully generate a theophylline-sensitive aptaSTAR.

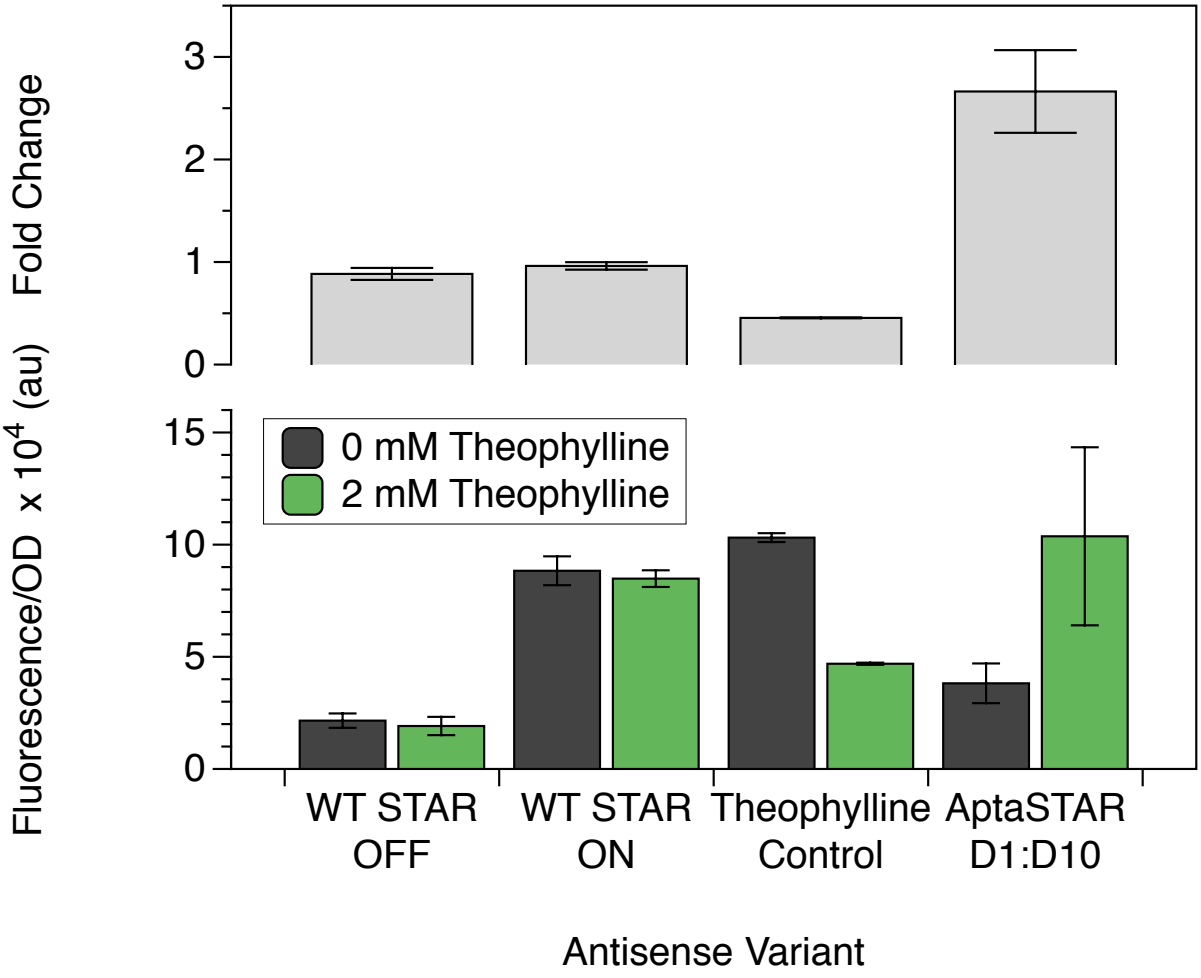


Figure 4.4 Theophylline sensitivity of an aptaSTAR variant. Upper graph shows fold change of fluorescence in response to 2 mM theophylline, while lower graph shows fluorescence values for tested aptaSTAR variant D1:D10 in comparison to the WT STAR system in its ON and OFF states and a theophylline-sensitive repressor.

4.3.3 Computational design of theophylline-responsive RNA activators

While the randomized screening approach succeeded in generating a theophylline-sensitive aptaSTAR based on the pbuE STAR, it yielded no hits for aptaSTARs based on the more highly-activating STARs like pAD1 and pT181. In an effort to generate more highly activating aptaSTARs, we turned once more to rational design. This time we pursued a new strategy based on computational prediction of probable RNA folding states in order to take into account the relative stabilities of the ON and OFF states and allow for proper switching (unlike my first attempts at rational design). In collaboration with Howard Salis of Penn State, we sought to engineer ligand sensitivity using thermodynamic calculations to predict the ideal sequence for our aptaSTARs. The approach had been previously validated in the design of synthetic riboswitches using calculated free energies of ribosome-binding to generate riboswitch variants predicted to correctly switch between ON and OFF states in response to a target ligand (Borujeni et al., 2015). In this case, we sought to generate aptaSTAR variants where the binding of theophylline would trigger STAR re-folding into a form favored to bind its sense target. Thermodynamic calculations were used to evaluate potential linker sequences (up to 30 nucleotides in length) to connect the STAR to the aptamer so as to maximize switching between the ON and OFF states in response to theophylline. The energy states under consideration in the design are schematized in Figure 4.5. The goal for generating potential aptaSTARs was to find designs predicted to have both a large free energy differences between the energy associated with anti-termination in the presence and absence of theophylline, in order to encourage switching

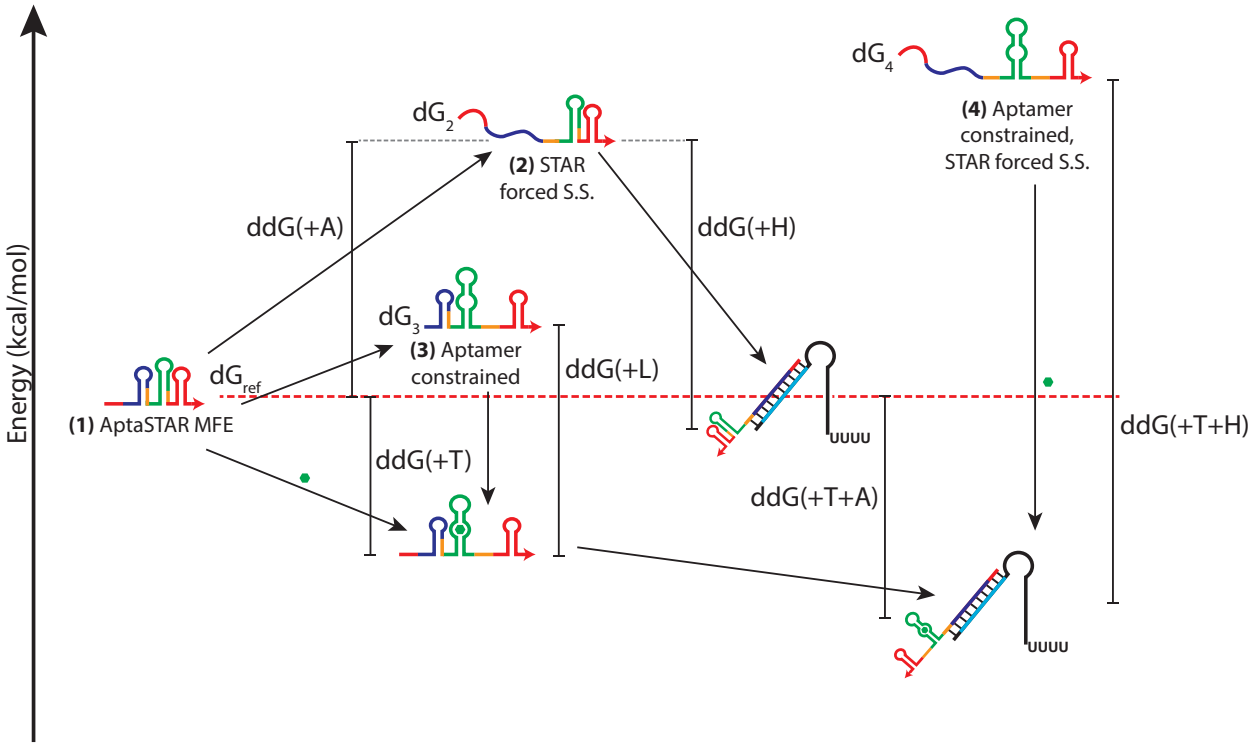


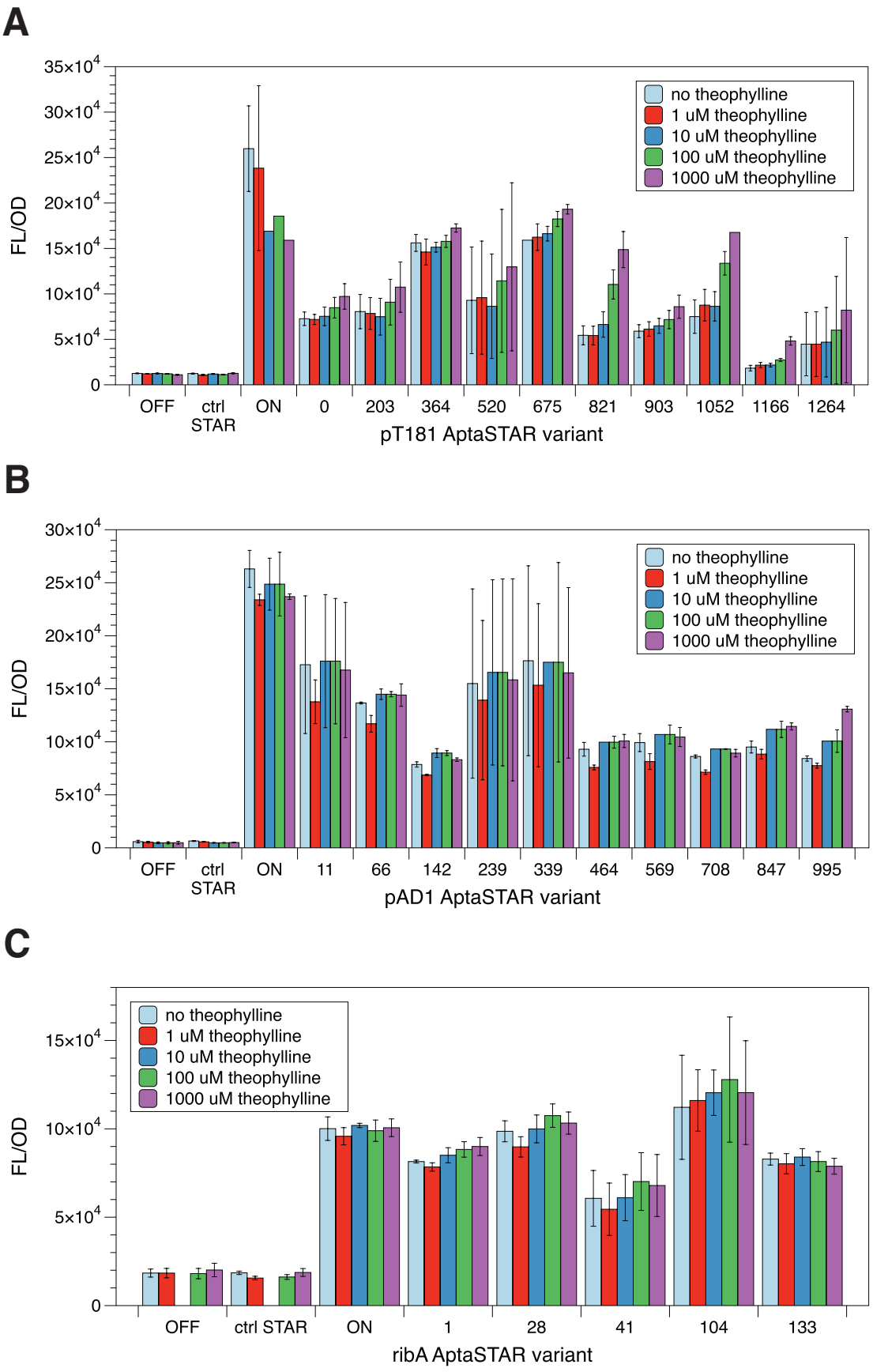
Figure 4.5 AptaSTAR energy states for computational design. State **(1)** represents the minimum free energy (MFE) fold for the unconstrained STAR with free energy dG_{ref}, while state **(2)** with energy dG_2 represents the aptaSTAR when the STAR anti-terminator portion is constrained to be single-stranded (S.S.). The aptaSTAR with the aptamer domain constrained to its ligand-binding fold is state **(3)**, with energy dG_3, and state **(4)** refers to the aptaSTAR with both STAR and aptamer constrained, yielding free energy dG_4. The free energy differences ddG(+A), ddG(+L), ddG(+T), and ddG(+H) refer respectively to the free energy changes associated with forming the anti-terminating STAR structure, ligand-binding, switching to the theophylline-bound state, and hybridizing to the target sense RNA. The computational algorithm generated designs that maximized the difference between ddG(+T+A) and ddG(+A) in order to encourage switching between states in response to theophylline.

between OFF and ON states, and thermodynamically favorable anti-termination in the presence of theophylline. The computational methods applied by Dr. Salis were used to generate large libraries of potential aptaSTAR designs based on the pT181, pAD1, and ribA STARs previously reported (Chappell et al., 2015).

From these libraries were picked a range of potential designs with equal energy gaps between anti-termination states with and without theophylline, but with varying ddG values and universally favorable (negative) ddG values for the energy of anti-termination in the presence of theophylline. The potential aptaSTARs were then cloned and tested *in vivo* with varying levels of theophylline from 0 μM to 1000 μM alongside the WT ON and OFF controls as well as a non-cognate STAR control (Figure 4.6). While the tested aptaSTARs based on the pAD1 and ribA STARs showed little theophylline responsiveness, with all variants activating SFGFP production regardless of theophylline concentration, a few of the designed pT181 variants did show theophylline sensitivity (Figure 4.6A), demonstrating increasing levels of activation in the presence of increasing levels of theophylline that were confirmed by follow-up experiment (Figure C.2), with the best variant (design 821) showing 3-fold activation in response to 1 mM theophylline.

Seeking to reach aptaSTAR activation levels more closely matching the greater than 50-fold activation seen by original pAD1 STAR, we generated another set of potential aptaSTARs where the computationally designed linker sequence was allowed to extend to 60 nucleotides (as opposed to 30 nucleotides previously). Our hypothesis was that a longer linker sequence would allow for more base-pair interactions to stabilize the interference of the STAR antisense—allowing for a lower OFF level and higher fold-activation. However, the resulting designs showed little theophylline sensitivity and high levels of activation in

Figure 4.6 Testing computationally designed aptaSTARs. AptaSTARs based on the pT181, pAD1, and ribA systems were tested for theophylline sensitivity *in vivo* with concentrations of theophylline ranging from 0 to 1000 μM alongside the WT type system in the OFF (no antisense) and ON (with STAR antisense) states and with a non-cognate STAR control.


A
35x10^4
30x10^4
25x10^4
20x10^4
15x10^4
10x10^4
5x10^4
0
FL/OD
no theophylline
1 uM theophylline
10 uM theophylline
100 uM theophylline
1000 uM theophylline
OFF
ctrl STAR
ON
0
203
364
520
675
821
903
1052
1166
1264
pT181 AptaSTAR variant
B
30x10^4
25x10^4
20x10^4
15x10^4
10x10^4
5x10^4
0
FL/OD
no theophylline
1 uM theophylline
10 uM theophylline
100 uM theophylline
1000 uM theophylline
OFF
ctrl STAR
ON
11
66
142
239
339
464
569
708
847
995
pAD1 AptaSTAR variant
C
15x10^4
10x10^4
5x10^4
0
FL/OD
no theophylline
1 uM theophylline
10 uM theophylline
100 uM theophylline
1000 uM theophylline
OFF
ctrl STAR
ON
1
28
41
104
133
ribA AptaSTAR variant

the absence of theophylline (Figure C.3), indicating that the linker sequence did not correctly interfere with the STAR antisense. Regardless, the previous results had validated the possibility of generating functional aptaSTARs via computation, although the uneven success of the designs indicates an imperfect model for predicting the energy states and the switching between them.

4.3.4 Alternate approaches: designing RNA activators responsive to MS2 coat protein and fluoride

Having shown the potential to generate theophylline-sensitive aptaSTARs, we turned our attention toward creating aptaSTARs able to sense alternative ligands in order to allow for multi-input circuits. The ligands I chose as targets were the fluoride ion and the phage MS2 coat protein, both of which have highly specific, well-characterized aptamers with sequence flexibility to allow for fusion engineering. The fluoride aptamer (Figure C.4), derived from a family of homologous bacterial riboswitches (Baker et al., 2012), has a bulge-stem structure with a conserved pseudoknot that generates a ligand-binding pocket where the fluoride ion is coordinated by a trio of hydrated magnesium ions (Ren et al., 2012). However, the bulge/loop portion of the aptamer show great variability between different riboswitch variants (Baker et al., 2012), offering an attractive opportunity for engineering alternate sequence interactions in an aptaSTAR context. The MS2 aptamer from the bacteriophage MS2 binds its target, the coat protein that forms the viral capsid, at the top of a stem-loop structure (Convery et al., 1998), allowing for sequence variation in the hairpin base in previously engineered variants (Qi et al., 2012).

To design the fluoride aptaSTAR, I used the *T. petrophila* variant of the fluoride aptamer, since its crystal structure has been solved in the context of the fluoride riboswitch (Ren et al., 2012), giving us good insight into which nucleotides and interactions are most important to maintain. I used a STAR variant based on the pAD1 system designed to bind its target via a toe-hold recognition region (unpublished data)—and fused it to the fluoride aptamer in a manner predicted to bind the STAR anti-terminator in the absence of aptamer stabilization by fluoride binding. However, when I tested its function *in vivo* against the parent pAD1 STAR, I found that showed no better response to fluoride than the parent STAR (Figure 4.7).

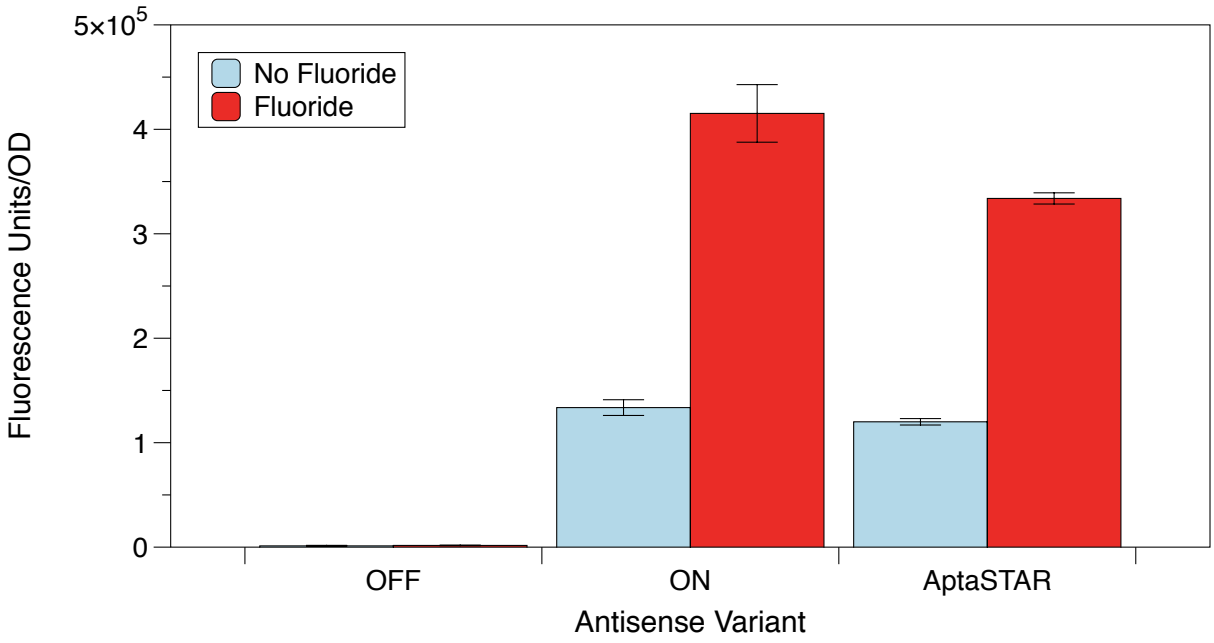


Figure 4.7 Fluoride aptaSTAR testing *in vivo*. Both the parent STAR and designed aptaSTAR appear to show fluoride-sensitive activation in response to fluoride—however, this is an artifact resulting from the greater SFGFP accumulation within the doubled growth time required to reach an equivalent OD in the presence of fluoride.

For the MS2 aptaSTARs, I used the nucleic acid design software NUPACK (Zadeh et al., 2010) to design the MS2 hairpin so that the 3' half of the hairpin would interfere with a portion of the fused STAR antisense while the 5' half of the hairpin varied in order to keep

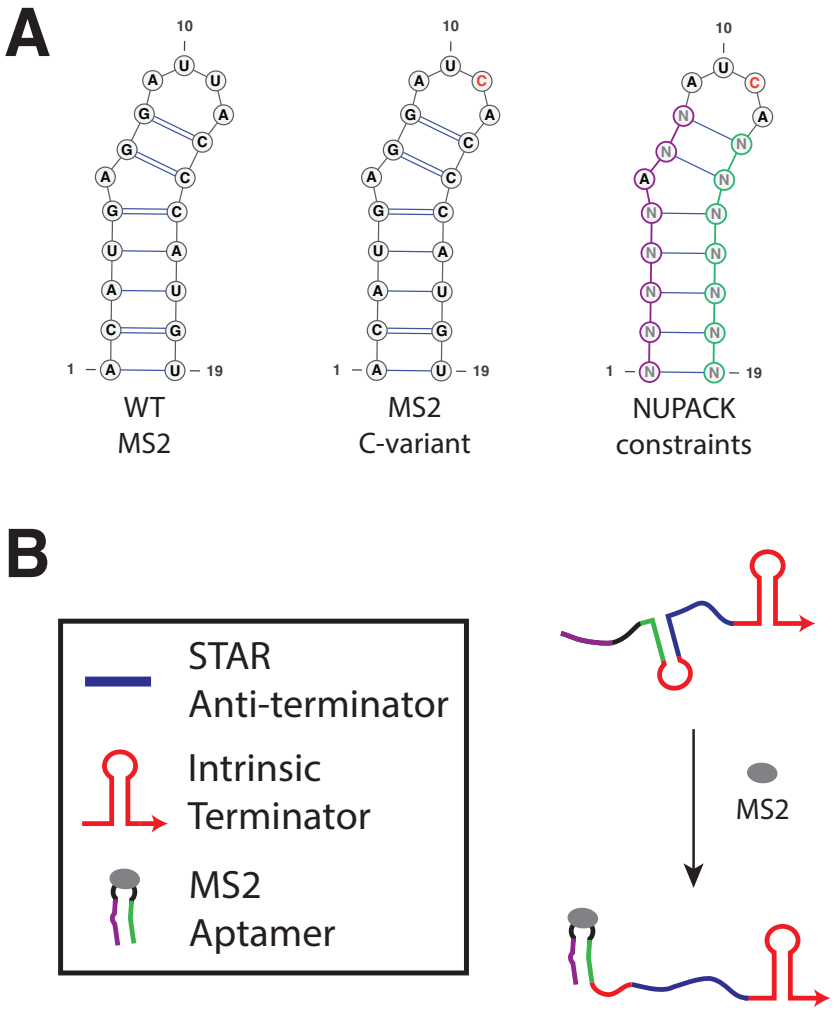


Figure 4.8 MS2 hairpin and aptaSTAR fusion design. **(A)** Structures of the wild-type MS2 hairpin, the MS2 C-variant previously reported to have stronger binding affinity (Convery et al., 1998), and the MS2 hairpin constraints used in NUPACK design of the MS2 aptaSTAR. **(B)** Designed mechanism of the MS2 aptaSTAR—the 3' side of the MS2 hairpin is designed to be complementary to a portion of the STAR antisense. MS2 protein addition should stabilize the MS2 aptamer, freeing the STAR antisense to activate transcription of its target.

sequence complementarity (Figure 4.8). The 'AUCA' tetraloop and the single-nucleotide adenine bulge on the 5' portion of the hairpin stem were maintained in order to keep MS2binding specificity, and the STAR antisense was the same pAD1 STAR variant used in the fluoride aptaSTAR design. Seven different MS2 aptaSTAR variants were designed using this strategy, each intended to interfere with a slightly different portion of the STAR antisense. The variants were then tested for activation *in vivo* in the presence and absence of the MS2 coat protein, produced from a separate plasmid via IPTG induction (Figure 4.9). However, none of the variants showed the appropriate lack of activation in the absence of MS2, instead showing activation on par with the control ON state regardless of whether MS2 was present. This lack of MS2 responsiveness indicated that the designed interference

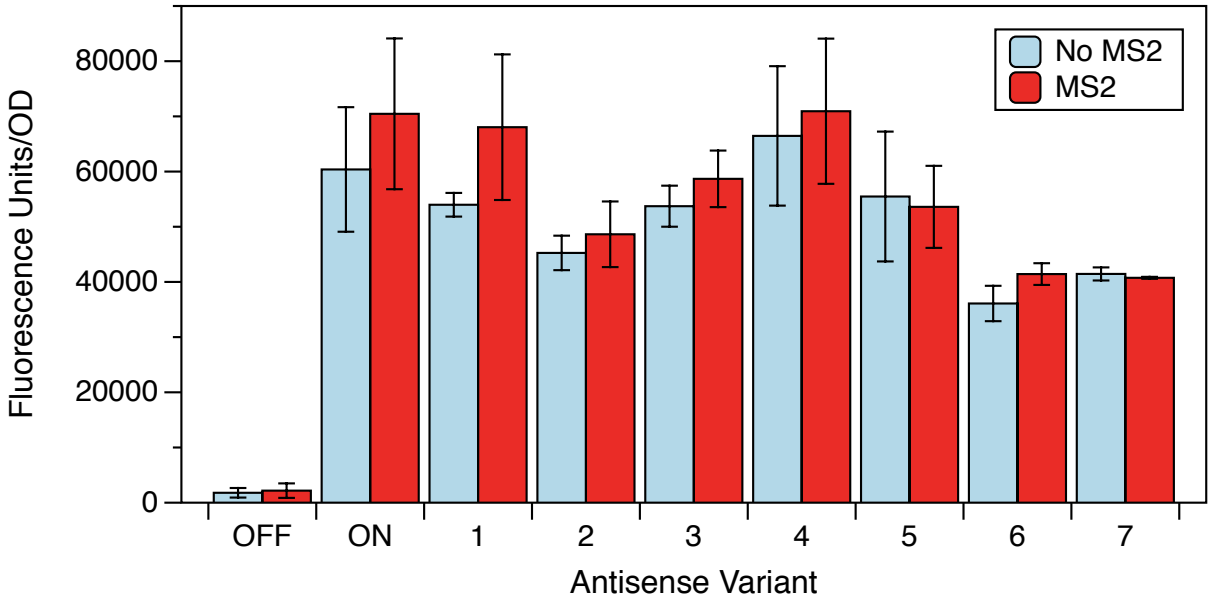


Figure 4.9 MS2 aptaSTAR testing *in vivo*. Testing of NUPACK-designed MS2 aptaSTAR variants for MS2-responsive activation alongside the basis pAD1 STAR system in its ON and OFF states indicated a lack of MS2 responsiveness rather than activation only in the presence of MS2.

hairpin may not have formed or may not have interfered with enough of the STAR antisense in order to successfully inhibit activation. Further exploration is required to pinpoint the reason for the observed lack of activation from this particular design strategy.

4.4 Conclusions

Despite the fact that not all design strategies were successful, our work resulted in the development of a handful of ligand-responsive aptaSTARs and has also demonstrated multiple feasible paths toward aptaSTAR generation, with both randomized screening and computational design succeeding in creating functional aptaSTARs with great potential applications in genetic circuitry. However, much research remains to be done in order to fully elucidate the underlying structural mechanism of aptaSTAR switching and to understand why designs predicted to be equally thermodynamically favorable have varying levels of theophylline sensitivity.

4.5 Materials and methods

Plasmid construction and cloning

STAR system plasmids followed the same architecture outlined in Chapter 3.3 and Appendix B, with the addition of aptamer and transducer region executed via iPCR, except in the case of the MS2 aptaSTAR plasmids, which were generated via Gibson Assembly.

Strains, media, and in vivo bulk fluorescence experiments

All cloning was performed in *E. coli* strain NEB Turbo (F' *proA*$^{+}$*B*$^{+}$ *lacI*^{q} *ΔlacZM15/ fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*Tet^{S} *endA1 thi-1 Δ(hsdS-mcrB)5)*,

while all bulk fluorescence experiments were performed in strain TG1 *(F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB))*, with the exception of the fluoride aptaSTAR experiments, which were performed in the CrcB ion transporter knock-out strain from the Keio Collection (Baba et al., 2006), based on strain BW25113 *F- Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) ΔcrcB763::kan λ- rph-1 Δ(rhaD-rhaB)568 hsdR514*. Media preparation and bulk fluorescence experiments were carried out as described in Chapter 3.3 with specific changes as described below.

Functional screening of theophylline aptaSTARs with randomized transducer regions

The functional screens associated with section 4.3.2 were conducted in the same manner as the other bulk fluorescence experiments from Chapter 3.3, but with duplicate 96-well blocks inoculated for the subculture step: one with chloramphenicol/carbenicillin M9 media containing 2 mM theophylline and one with chloramphenicol/carbenicillin M9 media with no theophylline. Each well-block screened also included one colony of the previously-reported theophylline-sensitive repressor (Qi et al., 2012) as a positive control in addition to the growth control and the parent STAR system in its ON and OFF states.

Functional testing of computationally designed theophylline aptaSTARs

The computationally designed aptaSTARs were tested as above, but in biological triplicate (three separate colonies for each tested pair of sense and antisense) with sub-culturing in chloramphenicol/carbenicillin M9 media containing 0 μM, 1 μM, 10 μM, 100 μM, and 1000 μM theophylline taking place in parallel. An additional non-cognate STAR OFF control

condition was also tested to confirm that transcription of an sRNA from a high-copy plasmid did not result in theophylline sensitivity.

Functional testing of fluoride and MS2 aptaSTARs

MS2 aptaSTAR testing took place using a three plasmid system: the p15a target plasmid, the colE1 aptaSTAR plasmid, and a BBR1 plasmid with lac-inducible expression of the MS2 coat protein required to bind to the MS2 aptamer. Two versions of the BBR1 plasmid were tested: one harboring a kanamycin resistance gene and the other encoding spectinomycin resistance. As a result, the solid and liquid media used had the appropriate triple antibiotics (carbenicillin and chloramphenicol as before and 100 μg/mL kanamycin or 50 μg/mL spectinomycin), and the *in vivo* testing in the with MS2 condition included induction with 1 mM IPTG during the sub-culture phase, while the condition without MS2 was not induced with IPTG. The remainder of the procedure followed the previously reported *in vivo* testing protocol from Chapter 3.

As mentioned previously, fluoride aptaSTAR testing took place in the BW25113 CrcB transporter knock-out from the Keio Collection in order to avoid having the added fluoride simply shuttled out of the cell. *In vivo* testing followed the protocol described previously, but with the addition of 0.5 mM NaF to one of the two sub-culture blocks. Moreover, because of the slow growth caused by the presence of fluoride, the cells grown in the presence of fluoride were grown for 8 hours and 40 minutes in order to match the optical density of the cells grown in the absence of fluoride.

4.6 Acknowledgements

Special thanks go to Howard Salis (Penn State) for his collaboration in the computational design of potential aptaSTARs and to Kyle Watters for his valuable advice in choosing potential aptamers.

4.7 Bibliography

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 1–11.

Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. Science *335*, 233–235.

Borujeni, A.E., Mishler, D.M., Wang, J., Huso, W., and Salis, H.M. (2015). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. Nucleic Acids Res. *44*, 1–13.

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013a). Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth. Biol. *2*, 463–472.

Ceres, P., Trausch, J.J., and Batey, R.T. (2013b). Engineering modular “ON” RNA switches using biological components. Nucleic Acids Res. *41*, 10449–10461.

Chappell, J., Takahashi, M.K., Meyer, S., Loughrey, D., Watters, K.E., and Lucks, J. (2013). The centrality of RNA for engineering gene expression. Biotechnol. J. *8*, 1379–1395.

Chappell, J., Takahashi, M.K., and Lucks, J.B. (2015). Creating small transcription activating

RNAs. Nat. Chem. Biol. *11*, 1–9.

Convery, M. a, Rowsell, S., Stonehouse, N.J., Ellington, a D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E., and Stockley, P.G. (1998). Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. Nat. Struct. Biol. *5*, 133–139.

Kim, P.B., Nelson, J.W., and Breaker, R.R. (2015). An Ancient Riboswitch Class in Bacteria Regulates Purine Biosynthesis and One-Carbon Metabolism. Mol. Cell *57*, 317–328.

Lynch, S.A., and Gallivan, J.P. (2009). A flow cytometry-based screen for synthetic riboswitches. Nucleic Acids Res. *37*, 184–192.

Lynch, S.A., Desai, S.K., Sajja, H.K., and Gallivan, J.P. (2007). A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. Chem. Biol. *14*, 173–184.

Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. *11*, 29–35.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science *306*, 275–279.

Meyer, S., Chappell, J., Sankar, S., Chew, R., and Lucks, J.B. (2016). Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. Biotechnol. Bioeng. *113*, 216–225.

Mishler, D.M., and Gallivan, J.P. (2014). A family of synthetic riboswitches adopts a kinetic trapping mechanism. Nucleic Acids Res. *42*, 6753–6761.

Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K., and Arkin, A.P. (2012). Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res. *40*,

5775–5786.

Ren, A., Rajashankar, K.R., and Patel, D.J. (2012). Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. Nature *486*, 85–89.

Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature *441*, 1167–1171.

Stojanovic, M.N., and Kolpashchikov, D.M. (2004). Modular aptameric sensors. J. Am. Chem. Soc. *126*, 9266–9270.

Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F., and Mörl, M. (2013). De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res. *41*, 2541–2551.

Winkler, W.C., and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. Annu. Rev. Microbiol. *59*, 487–517.

Zadeh, J.N., Wolfe, B.R., and Pierce, N.A. (2010). Nucleic Acid Sequence Design via Efficient Ensemble Defect Optimization. J. Comput. Chem. *32*, 439–451.

Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P., and Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nat. Struct. Biol. *4*, 644–649.

CHAPTER 5

CHARACTERIZING THE STRUCTURE-FUNCTION RELATIONSHIP OF A NATURALLY-OCCURRING RNA THERMOMETER

5.1 Abstract

A wide number of bacteria have been found to govern virulence and heat/cold shock responses using temperature-sensing RNAs known as RNA thermometers. A prime example is the *agsA* thermometer known to translationally regulate the production of the AgsA heat-shock protein in *Salmonella* using a "fourU" structural motif. Using the SHAPE-Seq RNA structure-probing method *in vivo* and *in vitro*, we both confirmed that the regulator functions by an equilibrium-determined partial melting of the helix containing the ribosome binding site and demonstrated the ability of SHAPE-Seq to detect ribosome binding to the *agsA* mRNA. These results display how subtle RNA structural changes can govern gene expression and illuminate the function of an important bacterial regulatory motif.

5.2 Introduction

Bacterial survival depends on a bacterium's ability to react appropriately to environmental cues, altering its behavior in response to its ever-changing surroundings. In particular, variations in temperature require coordinated system-wide responses in order to avoid deleterious effects (Kortmann and Narberhaus, 2012), resulting in the evolution of a wide array of molecular thermosensors—including RNA thermometers (Shapiro and Cowen, 2012). These RNA thermometers allow bacteria to govern wide-spread responses

to heat and cold shock and to control temperature-dependent virulence effects, an important consideration for pathogens with warm-blooded hosts (Konkel and Tilly, 2000).

The RNA thermometers regulating bacterial cold shock are perhaps the least well understood: although the details are yet to be elucidated, they are known to work by modulating temperature-dependent changes in transcript structure and stability for CspA mRNAs (Giuliodori et al., 2010; Yamanaka et al., 1999). This governs production of the CspA family of cold shock proteins, which act as RNA chaperones and transcription factors in conditions of cold stress (Uppal et al., 2008), thus regulating gene expression on a post-transcriptional level in response to environmental temperature.

Similarly, the thermosensing RNAs controlling heat shock also function post-transcriptionally— generally relying on the formation of temperature-sensitive secondary structure to fully or partially occlude the RBS (Narberhaus, 2010; Shapiro and Cowen, 2012), thus decreasing translation efficiency (de Smit and van Duin, 1990) in a temperature-dependent manner. The most widely characterized RNA thermometers are the broad family of ROSE (repression of heat shock gene expression) elements present in the 5' UTRs of heat shock genes (Kortmann and Narberhaus, 2012). First discovered in *Bradyrhizobium japonicum* (Narberhaus et al., 1998) and other rhizobia (Nocker et al., 2001), ROSE elements were later found to be common among α- and γ-proteobacteria, including *E. coli* (Waldminghaus et al., 2005, 2009), and amenable to identification through bioinformatics-based approaches (Waldminghaus et al., 2005, 2007a). ROSE elements (Figure 5.1) share a common sequence motif predicted to pair with the Shine-Dalgarno sequence in a hairpin (Narberhaus, 2010). Detailed structural information obtained through NMR spectroscopy indicates that this conserved ROSE motif depends on a non-

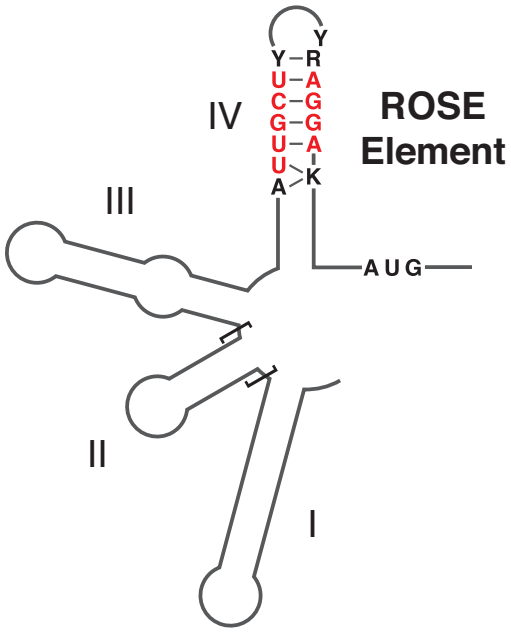


Figure 5.1 Consensus structure of repression of the heat shock gene expression (ROSE) element. Hairpins I, III, and the optional hairpin II vary widely in sequence and length, and the Shine-Dalgarno sequence is paired to a conserved UUGCU sequence within hairpin IV.

canonical interaction and an unusual base triple for hairpin stability and melting (Chowdhury et al., 2006).

Other heat shock RNA thermometers, like the one governing translation of the σ^{32} heat shock-dependent σ factor, involve the formation of complex structures that block ribosome entry without directly pairing with the RBS (Morita et al., 1999a, 1999b). Yet another mechanism, found in the RNA thermometer regulating the Hsp17 heat shock protein in cyanobacteria *Synechocystis* (Kortmann et al., 2011), depends upon an internal hairpin bulge to allow for stem melting in response to increased temperature (Wagner et

al., 2015). This same hairpin bulge pattern reappears in the *Salmonella enterica agsA* RNA thermometer (Figure 5.2A), where an internal stem bulge followed by a stretch of four uridines paired to the Shine-Dalgarno sequence confers thermosensing capability to the 5' UTR of the *agsA* gene encoding the aggregation suppression protein agsA (Waldminghaus et al., 2007b), a molecular chaperone that forms part of the heat shock response in *Salmonella* (Tomoyasu et al., 2003). This "fourU" motif has since been characterized *in vitro* via NMR spectroscopy, giving key insights into relative stabilities of the different pairings within the thermometer hairpin (Rinnenthal et al., 2010) and the magnesium-dependence of hairpin melting (Rinnenthal et al., 2011).

Although the fourU RNA thermometer motif was first coined in reference to the *agsA* thermometer from *Salmonella* (Waldminghaus et al., 2007b), the previously reported virulence control mechanism of the plague bacterium *Yersinia pestis* also relies a four uridine base-pairing to the Shine-Dalgarno sequence to grant temperature-dependent translational control (Hoe and Goguen, 1993). Similarly, the temperature-dependent virulence response of *Vibrio cholerae* was recently tied to a fourU element within the 5' UTR of ToxT, a key transcriptional activator of virulence factors (Weber et al., 2014). While other temperature-based virulence controls rely on alternate mechanisms (Johansson et al., 2002; Loh et al., 2009), the reoccurrence of the fourU motif in regulating both heat shock and virulence underlines its widespread utility as an RNA thermometer.

Here we seek to better understand the mechanistic details of its function *in vivo* by applying the SHAPE-Seq method of RNA structural determination to the fourU element of the *agsA* RNA thermometer. By combining structural probing data obtained under different temperatures and conditions we dissect the structural mechanism of this naturally

occurring RNA thermometer and highlight the effect of ribosome binding upon the structure of the *agsA* 5' UTR.

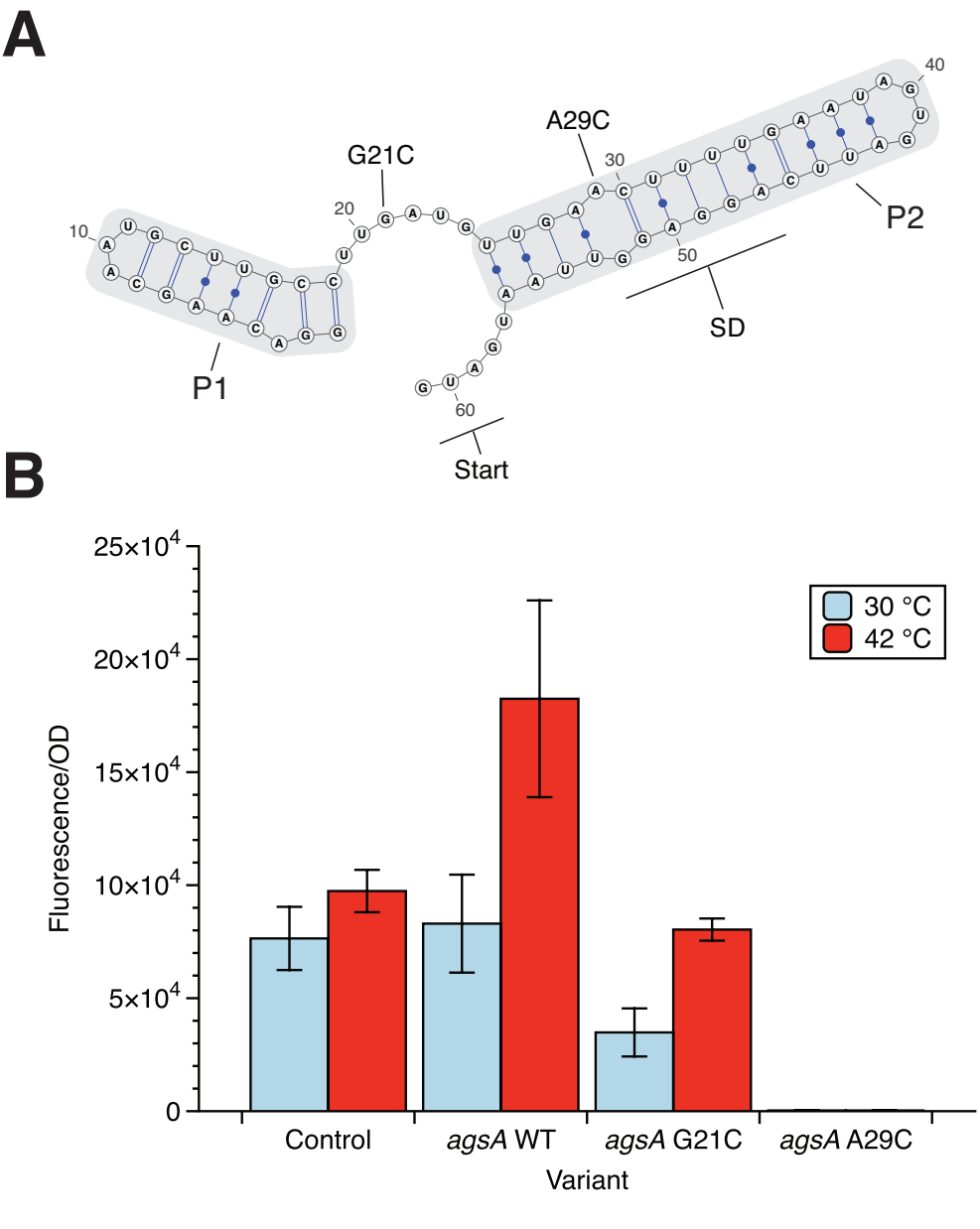
5.3 Results and discussion

Confirming agsA thermometer function in vivo

As a first step, we confirmed the temperature-sensitivity of the *agsA* thermometer by fusing the *agsA* 5' UTR directly to the super-folder green fluorescent protein (SFGFP) coding sequence and testing the construct's heat-responsiveness *in vivo* in *Escherichia coli* (Figure 5.2). Alongside the wild-type (WT) *agsA* thermometer, we also tested a control 5' UTR known to allow downstream translation at all temperatures and a pair of *agsA* mutants known to exhibit altered temperature responsiveness (Waldminghaus et al., 2007b). Specifically, the G21C mutation is predicted to extend the *agsA* P2 helix to include the AUG start codon at its base, resulting in tighter repression of translation at 30 °C than the wild-type. The A29C mutation removes a predicted single base-pair internal hairpin loop, thus strengthening the P2 helix and preventing RBS access at both 30 °C and 42 °C (Figure 5.2A).

To test the responsiveness of the constructs to heat-shock, we transformed them into *E. coli* and grew triplicate colonies overnight in liquid media. Duplicate subcultures of each colony were grown at 30 °C for 4 hours, and then one was subjected to a 30-minute heat-shock at 42 °C while the other remained at 30 °C. We then measured fluorescence and optical density via plate-reader, giving the normalized fluorescence data presented in Figure 5.2B. The control showed a slight increase in normalized fluorescence under heat-shock, corresponding to the faster production and maturation of SFGFP at the higher

Figure 5.2 Structure and function of the *agsA* thermometer. **(A)** Putative structure of the *agsA* 5' UTR RNA thermometer from *Salmonella* with key P1 and P2 helices, the Shine-Dalgarno (SD) sequence, start codon, and mutations of interest labeled. **(B)** Functional *in vivo* testing of the *agsA* thermometer and mutants using a fluorescent reporter gene demonstrated their responsiveness to heat-shock. The G21C mutation extends the P2 helix to include the start codon, resulting in lower fluorescence in both the low temperature and heat shock conditions. The A29C mutation closes the inner-hairpin loop within the P2 helix, causing complete repression of gene expression under both temperature conditions.


A
10
20
30
40
50
60
G21C
A29C
P1
P2
SD
Start
B
25×10^4
20×10^4
15×10^4
10×10^4
5×10^4
0
Fluorescence/OD
30 °C
42 °C
Control
agsA WT
agsA G21C
agsA A29C
Variant

temperature, but the WT *agsA* variant showed clear activation in response to the heat-shock, with a 2.5-fold increase in normalized fluorescence over the 30 °C condition. The G21C *agsA* variant gave lower overall levels of fluorescence but still showed increased gene expression in response to heat-shock, while the A29C variant did not express SFGFP at either 30 °C or 42 °C (Figure 5.2B). These results confirm the behavior of the *agsA* thermometer and the designed mutants as originally reported in Walminghaus et al. (2007b) and support the model of thermometer regulation tied to hairpin melting at higher temperatures, since mutations that stabilized the hairpin resulted in decreased gene expression.

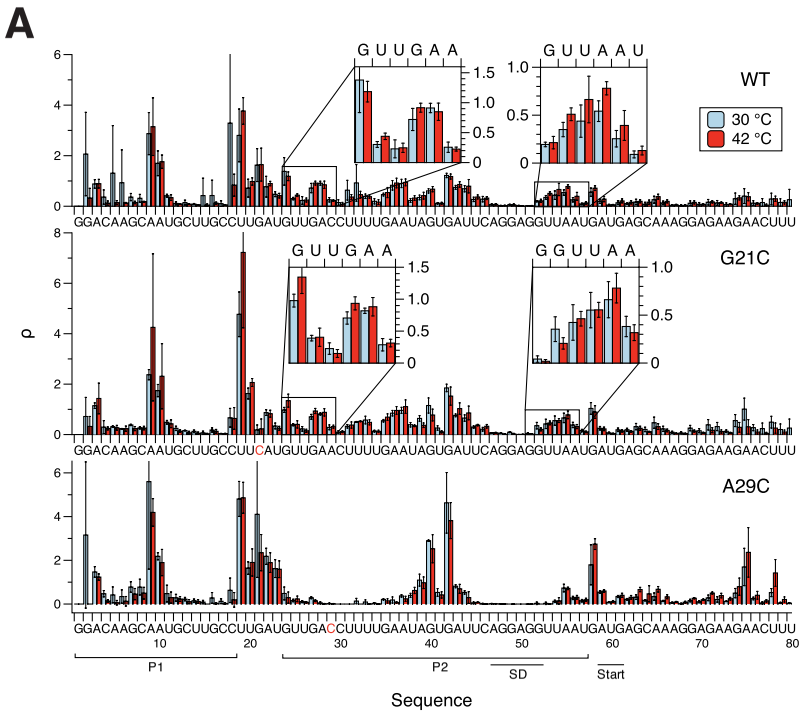
Determining agsA thermometer structure in vivo with SHAPE-Seq

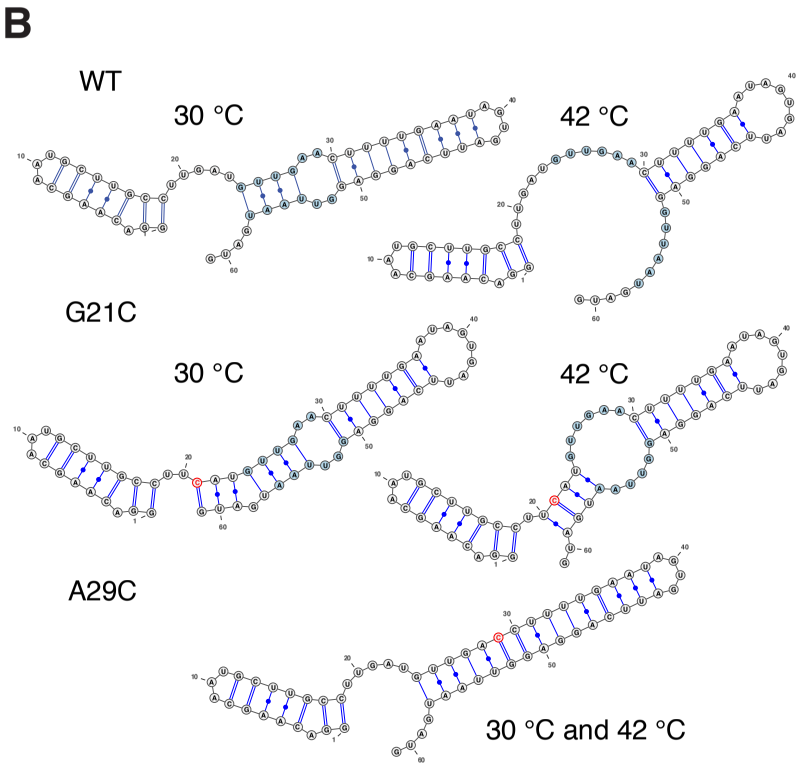
In order to determine the structure of the *agsA* 5' UTR *in vivo*, we turned to SHAPE-Seq, an RNA structure probing technique using a small chemical probe (1-methyl-7-nitroisatoic anhydride) that covalently modifies RNA nucleotides in a structure-dependent fashion (Lucks et al., 2011; Watters et al., 2015a, 2015b). The probe, which can penetrate through cell walls and membranes for in-cell structure determination, preferentially reacts with unconstrained nucleotides like those in loops and single stranded regions, but does not modify nucleotides constrained by base-pairing or stacking interactions (Bindewald et al., 2011). The locations of modifications are detected through reverse transcription and sequencing, allowing for the determination of the relative reactivities of the underlying RNA nucleotides to the SHAPE-Seq probe—high reactivities correspond to unpaired or unconstrained nucleotides, while low reactivities indicate base-pairing or constrained stacking interactions. These experimentally measured SHAPE reactivities can be

incorporated into structural prediction programs as a pseudo-energy constraint that favors structures matching the observed reactivity patterns in order to arrive at highly accurate predictions of RNA secondary structure (Loughrey et al., 2014; Low and Weeks, 2010; Reuter and Mathews, 2010).

We applied in-cell SHAPE-Seq to the *agsA* thermometer variants, probing the thermometer structures at 30 °C and 42 °C simultaneously with the bulk fluorescence assays presented in Figure 5.2. The resulting reactivity profiles (Figure 5.3A) were used to constrain RNA structural prediction (Figure 5.3B) using the RNAStructure software package (Reuter and Mathews, 2010) to predict RNA folding at the correct temperature. A comparison of the reactivity profiles and most likely RNA structures—the predicted minimum free energy (MFE) folds—between *agsA* variants and temperatures reveals a number of interesting differences. For the wild-type *agsA* RNA, slight reactivity increases in the nucleotides comprising the base of the P2 helix under heat-shock correspond to a predicted partial melting of the P2 hairpin at 42 °C. This correlates well with the functional data showing increased gene expression at higher temperatures, potentially mediated by better ribosome access to the Shine-Dalgarno region as the hairpin partially melts. While the G21C mutation does extend the P2 helix to contain the start codon at 30 °C, slight reactivity increases at 42 °C match a predicted expansion of the internal hairpin loop, correlating with the reported increase in fluorescence under heat-shock conditions. Neither of these temperature transitions was correctly predicted by RNAStructure in the absence of SHAPE constraints (Figure D.1), but the subtle reactivity changes measured by SHAPE-Seq influenced the folding algorithm enough to alter the predicted structures for the wild-type and the G21C variant. In comparison, the A29C mutation that closes the

Figure 5.3 SHAPE-Seq characterization of the *agsA* thermometer *in vivo.* **(A)** SHAPE-Seq normalized reactivity plots for *agsA* WT, G21C, and A29C variants from triplicate experiments *in vivo* performed at 30 °C and under 42 °C heat-shock conditions. **(B)** Experimentally-derived minimum free energy (MFE) RNA structures as predicted using the RNAStructure program with the reactivity values from (A) taken as pseudoenergy folding constraints. Blue-highlighted nucleotides undergo key structural changes according to the folding models and correspond to the inset boxes in (A), while red nucleotides denote variant mutations.


A
WT
30 °C
42 °C
G21C
A29C
ρ
GGACAAGCAAUGCUUGCCUUGAUGUUGACCUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
GGACAAGCAAUGCUUGCCUUCAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
GGACAAGCAAUGCUUGCCUUGAUGUUGACCUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
P1
P2
SD
Start
Sequence


B
WT
30 °C
42 °C
G21C
30 °C
42 °C
A29C
30 °C and 42 °C

inner hairpin loop creates a variant with strikingly low reactivities throughout the P2 helix and displays no predicted structural change between 30 °C and 42 °C, retaining an extended hairpin form even under heat-shock. This lack of hairpin opening corresponds to undetectable levels of gene expression at both temperatures, bolstering the proposed mechanism of translational control through hairpin opening and ribosome access.

However, it is worth noting that even at 42 °C, the most likely WT structure is still predicted to have perfect base-pairing of the Shine-Dalgarno sequence, somewhat at odds with the model of increasing ribosome access along with hairpin melting. Careful inspection of the RNAStructure ensemble base-pairing probabilities for the WT variant at

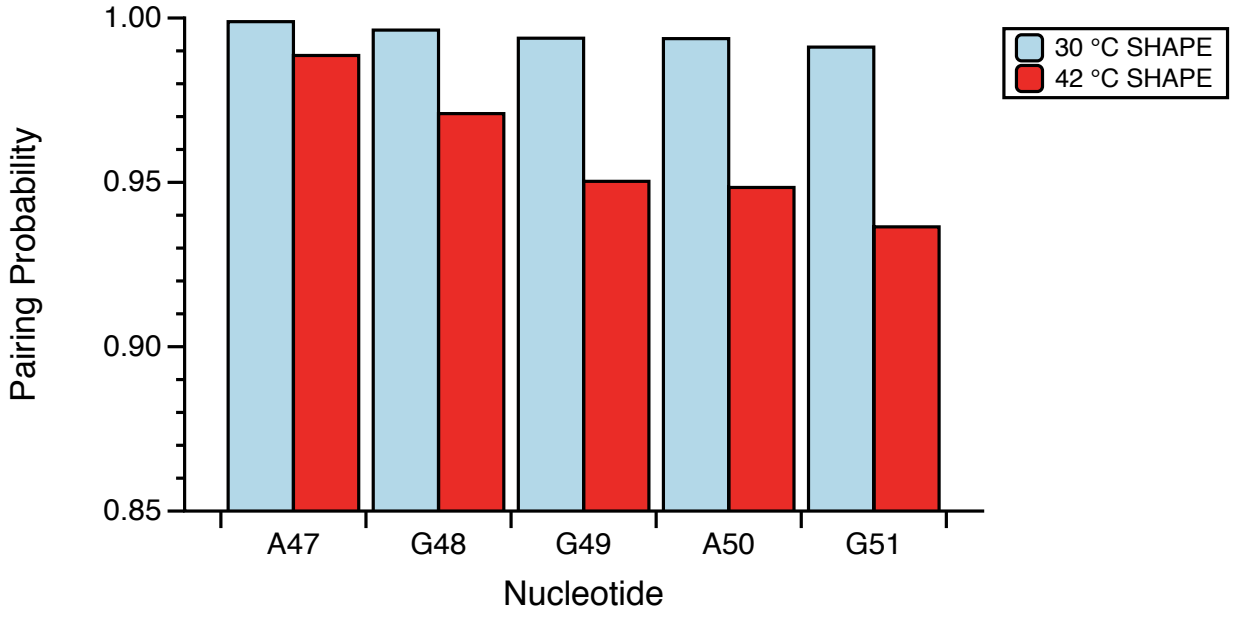


Figure 5.4 Base-pairing probabilities of the *agsA* WT Shine-Dalgarno sequence. The base-pairing probabilities for the first five nucleotides of the Shine-Dalgarno sequence as determined by the structural ensemble predictions constrained by *in vivo* SHAPE-Seq data.

30 °C and 42 °C (Figures D.2 and D.3), shows shifts toward lower probability base-pairing in the P2 stem at the higher temperature. However, the changes in the Shine-Dalgarno sequence are quite slight (Figure 5.4), with the ensemble-pairing probabilities shifting by just a few percentage points from approximately 99% pairing to 94-97% pairing, perhaps corresponding to a small population of transiently "breathing" hairpins. Intriguingly, this small ensemble shift in RBS accessibility is enough to more than double the expression of the downstream reporter gene, demonstrating how subtle shifts in RNA structural ensembles can lead to large changes in functional output.

Establishing the mechanism of the agsA thermometer's structural transition

Having observed *agsA* thermometer structure and function within the complex cellular environment where transcription and translation happen simultaneously, we turned to *in vitro* studies in order to decouple the two processes and better discern the underlying mechanism behind the shift in the *agsA* structural ensemble. We carried out our assays in the cell-free PURExpress protein synthesis system, first establishing that the *agsA* thermometer variants gave the expected reporter output in PURExpress from plasmid input, allowing for both transcription and translation (Supplementary Figure D.4). Next, we investigated whether the shift in the *agsA* structural ensemble depended on the kinetics of RNA folding during transcription (co-transcriptional folding) or the equilibrium folding of already transcribed RNAs. To do so, we tracked fluorescent reporter gene expression via plate reader for PURExpress reactions containing equal amounts of pre-transcribed, purified mRNAs for the control and *agsA* variants, thus eliminating any effect from co-transcriptional folding at a particular temperature. The resulting hour-long fluorescence

traces (Figure 5.5A) recapitulated the function observed *in vivo*, demonstrating higher fluorescence relative to the control at 42 °C than 30 °C for the WT and G21C mutants and consistently low fluorescence for the A29C variant. The baseline increase in gene expression of the control from 30 °C to 42 °C is higher than seen *in vivo*, due partly to the longer duration of the heat shock and partly to the magnifying effect of the high-yield PURExpress system. Figure 5.5B presents the measured rates of fluorescence increase during linear growth, reflecting the same trends observed in overall fluorescence. The

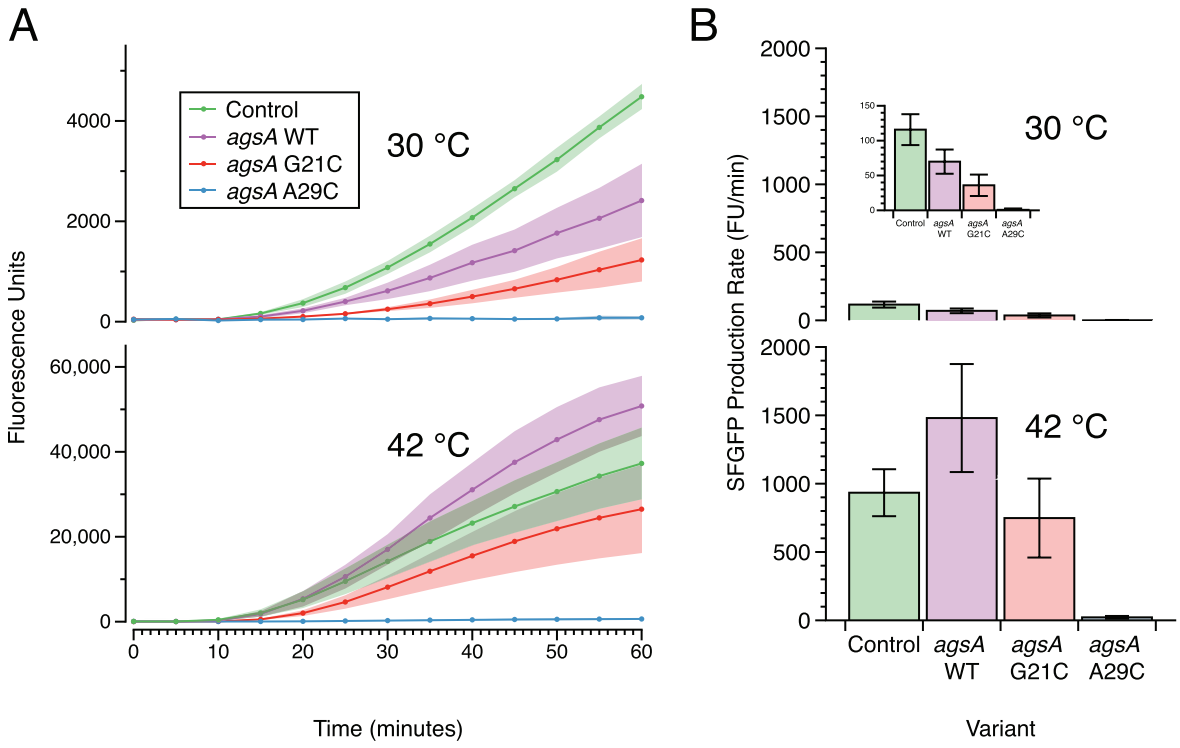
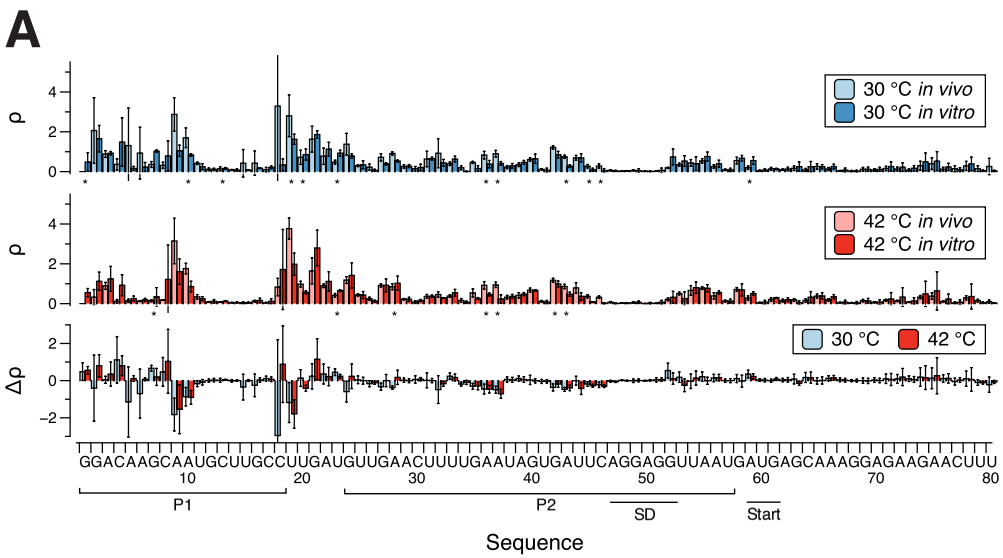
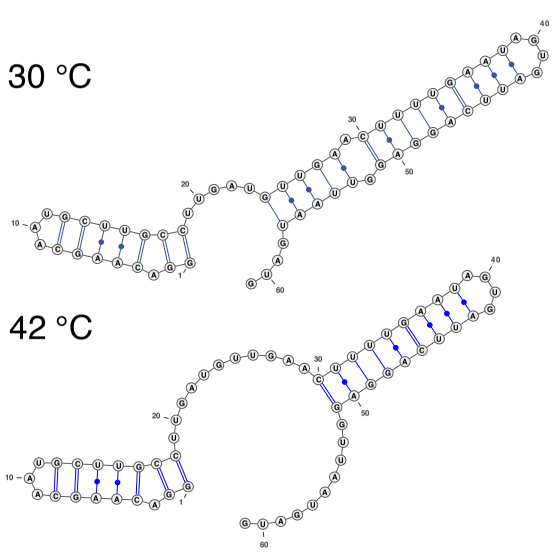


Figure 5.5 Testing *agsA* thermometer function in a cell-free protein synthesis system using purified pre-transcribed mRNA. **(A)** Fluorescence growth curves for translation of SFGFP from *agsA* construct and control mRNA in the PURExpress protein synthesis system at 30 °C and 42 °C. **(B)** SFGFP production rates during the linear synthesis regime for *agsA* constructs at 30 °C (45-50 minutes) and 42 °C (30-35 minutes).

Figure 5.6 SHAPE-Seq characterization of the *agsA* thermometer *in vitro* in PURExpress. **(A)** SHAPE-Seq normalized reactivity plots for *agsA* WT mRNA from triplicate experiments *in vitro* performed on mRNAs at 30 °C and 42 °C, graphed alongside previous *in vivo* reactivities for comparison. Stars denote statistically significant differences as determined by heteroscedastic Welch's T-test, while the difference plot shows the magnitude of reactivity change (*in vitro* reactitivies minus *in vivo* reactivities). **(B)** Experimentally-derived RNA structures as predicted using the RNAStructure program with the *in vitro* reactivity values from (A) taken as pseudoenergy folding constraints.


A
ρ
30 °C in vivo
30 °C in vitro
42 °C in vivo
42 °C in vitro
Δρ
30 °C
42 °C
GGACAAGCAAUGCUUGCCUUGAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
P1
P2
SD
Start
Sequence

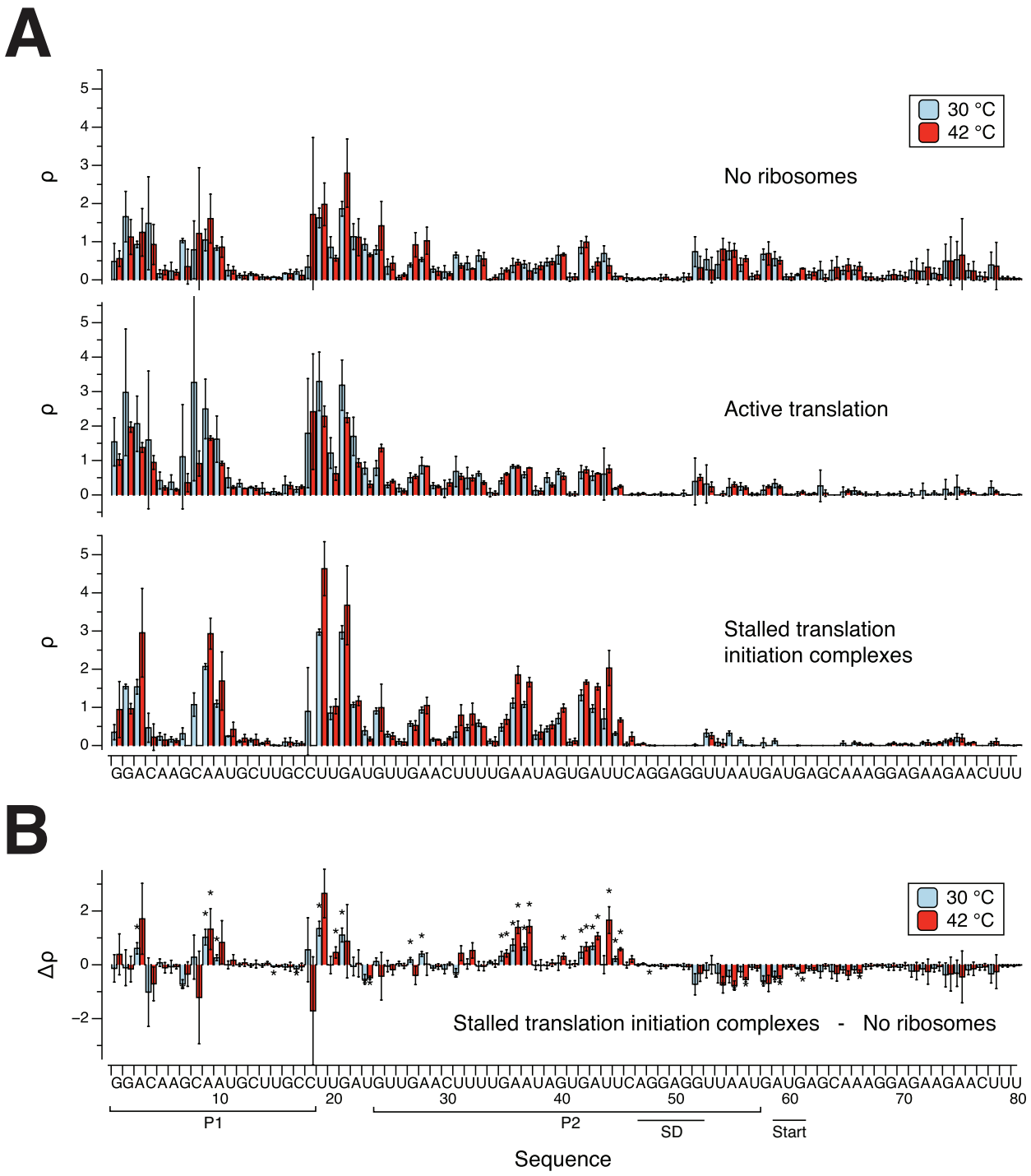

B
30 °C
42 °C

temperature responsiveness of the pre-transcribed *agsA* construct mRNAs demonstrates that RNA thermometer function does indeed depend on thermodynamic equilibrium folding as previously suggested by Waldminghaus et al. (2007b), as opposed to relying on a co-transcriptional effect where the kinetics of either transcription or simultaneous transcription and translation determine RNA structure or RBS accessibility.

Investigating agsA thermometer folding in vitro using SHAPE-Seq

Having confirmed the basic mechanism of *agsA* thermometer function, we set about investigating *agsA* thermometer folding *in vitro* to explore whether we could detect a stronger signal for structural switching in the absence of other cellular factors. Performing SHAPE-Seq *in vitro* (in the PURExpress cell-free system without ribosomes) on *agsA* WT mRNA at 30 °C and 42 °C produced the reactivity plots presented in Figure 5.6A (plotted alongside *in vivo* reactivities for comparison). Quite notably, the reactivity profiles at both temperatures are markedly similar, with only a handful of nucleotides showing statistically significant differences in reactivity (labeled with stars). When the *in vitro* reactivities are used to constrain the RNA structural prediction, the resulting structural predictions (Figure 5.6B) parallel those seen in under *in vivo* conditions, but with two key differences: at 30 °C the internal loop in the P2 helix is predicted to be one base-pair smaller, and at 42 °C the P2 stem loop is predicted to be two base-pairs smaller. Similar patterns in reactivity and structure were seen for *in vitro* experiments conducted in buffer as opposed to PURExpress, demonstrating a stable pattern of helix formation with slight variations in loop size. (See Figures D.5 and D.6.)

Figure 5.7 Investigating the effect of translation on *agsA* WT structure and reactivity. **(A)** SHAPE-Seq reactivity plots for triplicate experiments carried out *in vitro* in PURExpress without ribosomes, with active translation, and with stalled translation initiation complexes generated by supplying only methionine. **(B)** Difference plot of reactivities with stalled translation initiation complexes minus those from the no ribosome case. Stars represent statistical significance.


A
ρ
30 °C
42 °C
No ribosomes
Active translation
Stalled translation
initiation complexes
GGACAAGCAAUGCUUGCCUUGAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
B
Δρ
Stalled translation initiation complexes - No ribosomes
10
20
30
40
50
60
70
80
P1
P2
SD
Start
Sequence

Determining the effect of translation and ribosome binding on agsA structure

Having established the baseline structure of the *agsA* thermometer in the PURExpress system without ribosomes to be very similar to the *in vivo* structure, we next set about determining the reactivity changes caused by ribosome binding and active translation in the clearly defined *in vitro* system. To do so we performed additional SHAPE-Seq experiments *in vitro* in PURExpress involving both active translation and stalled translation initiation complexes (TIC) (Figure 5.7). For the active translation experiments, we probed PURExpress reactions assembled with ribosomes, and for the stalled translation initiation experiments, we probed PURExpress reactions assembled lacking all amino acids except for methionine—thus causing the translation initiation complex to stall after the addition of the first amino acid. The major reactivity changes caused by the addition of stalled translation initiation complexes become clear when the reactivity spectra (Figure 5.7A) are compared—the magnitudes of the reactivity differences are plotted in Figure 5.7B. There is a stark reactivity drop directly downstream of the ribosome binding site, corresponding to the ribosome footprint as the stalled translation initiation complex blocks SHAPE modification. Moreover, many of the nucleotides that formed the P2 helix, on the 5' side and the upper portion of the 3' side, undergo large increases in reactivity corresponding to increased flexibility brought about by the dissolution of the P2 helix in concert with ribosome binding. Larger changes in reactivity at 42 °C than 30 °C suggest greater ribosome access at the higher temperature, matching the functional data seen *in vivo* and *in vitro*. However, given that the ribosomes were supplied in excess to the mRNA (8:1 ratio) and allowed 20 minutes to reach binding equilibrium, it is unsurprising that the reactivities indicate ribosome binding to a large portion of the mRNA ensemble under both

temperature conditions—the presence of background gene expression at 30 °C in functional testing indicates some level of ribosome access even at the lower temperature, most likely linked to a small fraction of the ensemble undergoing transient hairpin breathing that allows for ribosome capture.

Comparing the reactivities for the no ribosome and stalled translation initiation complex cases aids with interpretation of the reactivity spectra from the active translation experiments. The reactivities immediately downstream of the ribosome binding site are lower than without ribosomes but higher than with stalled translation initiation complexes, as one would expect for an ensemble measurement in which some but not all of the mRNAs have ribosomes blocking the initiation site or the beginning of the downstream gene. That the reactivities are slightly lower than *in vivo* (Figure D.7) may indicate a higher level of gene expression within PURExpress without the need to compete with other mRNAs for translational machinery.

Discussion of agsA structural mechanism and perspectives

Combined with the structural and functional data from *in vivo* experiments, the PURExpress probing data confirms the previously suggested mechanism of *agsA* thermometer action: higher temperatures lead to partial melting of the hairpin containing the ribosome binding site, thus allowing for translation initiation. Moreover, we show clear indications that the mechanism operates in thermodynamic equilibrium as opposed to depending upon co-transcriptional RNA folding. In addition, we have shown strong evidence for *agsA* structural switching *in vivo*, unlike previous structural studies of fourU thermometers (Rinnenthal et al., 2010, 2011; Waldminghaus et al., 2007b), which took

place in vitro and in one case, in non-physiologically relevant conditions (Rinnenthal et al., 2010)—NMR buffer in the absence of magnesium, a key determinant of RNA folding and structure (Misra and Draper, 2002). Subtle differences between the RNA structures observed *in vivo* and *in vitro* reaffirm the importance of probing RNAs in their native environment, although *in vitro* assays allowed for the dissection of important mechanistic details and demonstration of the opening of the *agsA* P2 helix in response to ribosome binding.

The elegantly simple hairpin melting mechanism behind *agsA* thermo-sensing offers a potential template for future design of synthetic RNA thermometers with different properties—for instance, alternate ranges of temperature sensing or levels of gene activation could be achieved by altering the structure of the thermometer hairpin. Previous work in engineering temperature-sensing RNAs has drawn on a number of design strategies, from computational methods (Neupert and Bock, 2009; Neupert et al., 2008; Waldminghaus et al., 2008) to rational design (Hoynes-O'Connor et al., 2015), with varying levels of success, but the simplicity and modularity of the fourU thermometer make it a particularly attractive starting point for future engineering. In fact, recent work by has drawn on these particular properties to generate combined riboswitch-RNA thermometer systems allowing for complex temperature and ligand based regulation of synthetic systems. It is our hope that a better understanding of the structural mechanism underlying the *agsA* fourU thermometer will promote future engineering of this fascinating molecular sensor.

5.4 Materials and methods

Cloning and plasmid construction

Relying on the sequences reported by Waldminghaus et al. (2007), the *agsA* 5' UTR in both its wild-type and point-mutant forms was cloned into a p15a plasmid backbone harboring chloramphenicol resistance. Using inverse PCR (iPCR) for scarless insertion, the *agsA* sequence was introduced directly between the *E. coli* consensus promoter and the SFGFP start codon. See Appendix D for plasmid maps (Figure D.8) and sequences (Table D.1).

In vivo bulk fluorescence assays and in-cell SHAPE-Seq

All bulk fluorescence and in-cell SHAPE-Seq experiments were performed in *Escherichia coli* (*E. coli*) strain TG1 *(F'traD36 lacIq Delta(lacZ) M15 pro A+B+/supE Delta(hsdM-mcrB)5 (rk- mk- McrB-) thi Delta(lac-proAB))* with three independent biological replicates. For each independent replicate, an empty control plasmid (JBL 001), gene expression control plasmid (JBL2348), and *agsA* WT and mutant plasmids were transformed into chemically competent *E. coli* TG1 cells, plated on Difco LB + Agar plates with 34 mg/ml chloramphenicol and incubated overnight at 37 °C for approximately 17 hours. After overnight incubation, the plates were removed from the incubator and kept at room temperature for approximately 7 hours. Three separate colonies were picked for each control and one for each tested variant, and each colony was used to inoculate 1 mL of LB media with 34 mg/mL chloramphenicol in a 2 mL 96-well block (Costar 3960). The block was covered with a breathable seal (Aeraseal BS-25) and incubated at 37 °C while shaking at a speed of 1,000 rpm in a Labnet Vortemp 56 bench-top shaker for 18 hours

overnight. From these overnight cultures 24 μL were taken and used to inoculate two duplicate 96-well blocks containing 1.2 mL LB media with 34 mg/mL chloramphenicol. The blocks were shaken at 30 °C and 1,000 rpm for 4 hours, and then one of the blocks was moved to shake at 42 °C (heat-shock condition), while the other was kept at 30 °C. From there, the in-cell SHAPE-Seq probing and the bulk fluorescence sample preparation were carried out as described by Watters et al. (2015a and 2015b), using 1 mL of the sample for in-cell SHAPE-Seq and 150 μL for the bulk fluorescence measurements. Any deviations from the published protocol are described below. The samples' optical density (OD) at 600 nm and fluorescence (485 nm excitation, 520 nm emission) were measured with a Biotek Synergy H1 plate reader for the bulk fluorescence. Following RNA probing as described (Watters et al., 2015b), the modified and unmodified RNAs were extracted using the TRIzol Max Bacterial RNA Isolation Kit (Thermo Fisher). The extracted RNAs were reverse transcribed from within the SFGFP gene (primer = CAACAAGAATTGGGACAACTCCAGTG) using 3 μL of 50 nM (not 0.5 μM) primer per reaction. The following RNA hydrolysis step was extended from 5 minutes to 15 minutes, and the partial neutralization with HCl was omitted. Subsequent ethanol precipitation, adaptor ligation, and bead purification steps followed the published protocol exactly (Watters et al., 2015b), and the quality analysis and library construction were performed following the previously reported PCR modifications to minimize side-product formation for libraries from weakly expressed RNAs (Watters et al., 2015a). Data analysis was executed using the Spats pipeline as described by Watters et al. (2015a) in order to generate raw θ reactivities (Aviran et al., 2011), which were then converted to ρ reactivities by multiplying by the length of the probed RNA (Loughrey et al., 2014) and normalized using the box-plot strategy described by Deigan et al. (2008).

mRNA preparation

DNA templates for *in vitro* transcription were generated using PCR to add the T7 promoter to the *agsA* variants and control and to amplify the region from the 5' UTR to the intrinsic terminator following the SFGFP gene. The resulting templates were transcribed as described in section 2.5, with triple the amount of T7 polymerase to correct for differences in activity among different preparations of the enzyme. The resulting mRNAs were purified using Ampure XP beads (Beckman Coulter) following the manufacturer's protocol and run alongside RiboRuler High Range RNA ladder (Thermo Fisher) on a 3% denaturing polyacrylamide gel. The gel was stained using SYBR Gold (Thermo Fisher) and visualized on a ChemiDoc XRS+ (Bio-Rad) to check for template length and purity.

PURExpress assays

The cell-free assays showing temperature-sensitive expression from both DNA templates and pre-transcribed mRNA were performed in the PURExpress ΔRibosome Kit (New England Biolabs). To each transcription-translation reaction were added 4 µL of kit Solution A, 1.2 uL Factor Mix, 1.8 uL *E. coli* ribosomes (New England Biolabs), 0.75 µL *E. coli* RNA polymerase holoenzyme (New England Biolabs), 0.4 uL SUPERase In RNase Inhibitor (Thermo Fisher), 100 ng plasmid template, and enough nuclease-free water to reach a total volume of 10 µL. Each translation-only reaction contained 4 µL Solution A, 1.2 uL Factor Mix, 1.8 uL *E. coli* ribosomes, 3.24 pmol purified mRNA template, and enough nuclease-free water to reach a total volume of 10 µL. Reactions were pipetted into a clear-bottomed, black 384-well plate (Griner) loaded onto a Biotek Synergy H1 plate reader pre-warmed to either 30 °C or 42 °C where fluorescence (485 nm excitation, 520 nm emission) was

monitored for 3 hours with measurements being taken every 5 minutes. Measurements were corrected for background fluorescence by subtracting the average reading for three empty wells.

In vitro SHAPE-Seq in buffer

In vitro SHAPE-Seq was carried out largely following the *in vitro* SHAPE-Seq protocol from Watters et al. (2015a) performed on 10 pmol of purified mRNA folded and probed at either 30 °C or 42 °C. After modification, the RNAs were ethanol precipitated with the addition of 90 µL nuclease-free water, 10 µL 3 M sodium acetate (pH 5.5), 1 µL glycogen, and 300 µL ice-cold 100% ethanol, followed by freezing for 30 minutes at -80 °C and centrifugation at 4 °C and 15,000 rpm for 30 minutes. All liquid was aspirated, and the pellets were washed with 200 µL 70% ethanol before further centrifugation for 2 minutes, removal of the wash liquid, and re-suspension in 10 µL of nuclease-free water. From there, the remainder of experimental the protocol—from reverse transcription through raw reactivity calculation followed the published method from Watters et al. (2015a). Reactivities were normalized as above.

In vitro SHAPE-Seq in PURExpress

For experiments probing the structure of the *agsA* WT and control mRNAs in PURExpress without ribosomes, with active translation, or with stalled translation initiation complexes, 3.24 picomoles of purified mRNA were probed at either 30 °C or 42 °C. The translation initiation complex reactions were assembled using the PURExpress Δaa ΔtRNA kit (New England Biolabs), with each 10 µL reaction containing 2 µL kit Solution A, 1

μL 3 mM L-methionine (Sigma Aldrich), 1 μL kit tRNA, 3 μL kit Solution B, and 3.24 picomoles of mRNA—the lack of amino acids other than methionine resulted in ribosomal complex stalling at the start codon. The PURExpress probing experiments without ribosomes were carried out in a similar fashion using the PURExpress ΔRibosome Kit (New England Biolabs), with each 10 μL reaction containing 4 μL kit Solution A, 1.2 μL Factor Mix, and 3.24 picomoles of mRNA; the PURExpress active translation experiments contained 1.2 uL of *E. coli* ribosomes in addition. Reactions were incubated for 20 minutes (or 35 minutes for active translation) at 30 °C or 42 °C before being split for modification—5 μL of reaction were added to 0.56 μL 65 mM 1M7 in DMSO (positive channel) and to 0.56 μL DMSO (negative channel) and incubated for an additional minute before being transferred to ice. From there, the modified and unmodified RNA were extracted using the Trizol Max Bacterial RNA Isolation Kit (Thermo Fisher), and the remainder of the protocol followed the same method as the in-cell SHAPE-Seq samples.

Statistical analysis and data presentation

All experiments were carried out with three distinct replicates. Error bars represent sample standard deviation, and significance was determined using two-sided heteroscedastic t-tests. For difference plots, error was propagated using standard propagation of error techniques.

RNAStructure structure prediction

Minimum free energy RNA structures were predicted using the fold command from the RNAStructure command-line interface (Reuter and Mathews, 2010). Box-plot

normalized reactivities were specified as pseudoenergy constraints using RNAStructure's "-sh" option, and the SHAPE restraint intercept (*b*) and slope (*m)* were set at -0.3 kcal/mol and 1.1 kcal/mol, respectively, based on the optimum values for weighting ρ reactivities as determined by Loughrey et al. (2014). Temperature was specified as either 303.15 K (30 °C) or 315.15K (42 °C) using the "-t" temperature option. Structures were output as connectivity (ct) files, converted to dot-bracket notation (dbn) using the ct2dot utility from RNAStructure, and drawn using the VARNA graphical user interface (Hofacker, 2003). Structures represent predictions for *agsA* variant 5' UTRs. Structures including the full mRNA sequence constrained by SHAPE data for the 5' UTR and SFGFP leader sequence are presented in Figures D.9 through D.12. Partition function generation of the RNA probability pairing matrices was performed using the RNAStructure webserver's partition function, with the same temperature, SHAPE constraints, and pseudoenergy constraints as the minimum free energy structure prediction.

5.5 Acknowledgements

Many acknowledgements go to Matt Verosloff for conducting in-house replication experiments, to James Chappell and Kyle Watters for fruitful discussions about experimental design and to the Jewett Lab at Northwestern University for providing purified ribosomal subunits to test a preliminary hypothesis.

5.6 Bibliography

Aviran, S., Trapnell, C., Lucks, J.B., Mortimer, S.A., Luo, S., Schroth, G.P., Doudna, J.A., Arkin, A.P., and Pachter, L. (2011). Modeling and automation of sequencing-based characterization of RNA structure. Proc. Natl. Acad. Sci. *108*, 11069–11074.

Bindewald, E., Wendeler, M., Legiewicz, M., Bona, M.K., Wang, Y., Pritt, M.J., Le Grice, S.F.J., and Shapiro, B. a (2011). Correlating SHAPE signatures with three-dimensional RNA structures. RNA *17*, 1688–1696.

Chowdhury, S., Maris, C., Allain, F.H.-T., and Narberhaus, F. (2006). Molecular basis for temperature sensing by an RNA thermometer. EMBO J. *25*, 2487–2497.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2008). Accurate SHAPE-directed RNA structure determination a SHAPE experiment can be interpreted as a pseudo-free energy high accuracy . Proc. Natl. Acad. Sci. *106*, 97–102.

Giuliodori, A.M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C.O., and Pon, C.L. (2010). The cspA mRNA Is a Thermosensor that Modulates Translation of the Cold-Shock Protein CspA. Mol. Cell *37*, 21–33.

Hoe, N.P., and Goguen, J.D. (1993). Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. J. Bacteriol. *175*, 7901–7909.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Hoynes-O'Connor, A., Hinman, K., Kirchner, L., and Moon, T.S. (2015). De novo design of heat-repressible RNA thermosensors in E. coli. Nucleic Acids Res. 1–14.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., and Cossart, P. (2002). An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell

110, 551–561.

Konkel, M.E., and Tilly, K. (2000). Temperature-regulated expression of bacterial virulence genes. Microbes Infect. *2*, 157–166.

Kortmann, J., and Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. Nat. Rev. Microbiol. *10*, 255–265.

Kortmann, J., Sczodrok, S., Rinnenthal, J., Schwalbe, H., and Narberhaus, F. (2011). Translation on demand by a simple RNA-based thermosensor. Nucleic Acids Res. *39*, 2855–2868.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. Cell *139*, 770–779.

Loughrey, D., Watters, K.E., Settle, A.H., and Lucks, J.B. (2014). SHAPE-Seq 2.0 : systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. *42*, 1–10.

Low, J.T., and Weeks, K.M. (2010). SHAPE-directed RNA secondary structure prediction. Methods *52*, 150–158.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Misra, V.K., and Draper, D.E. (2002). The linkage between magnesium binding and RNA folding. J. Mol. Biol. *317*, 507–521.

Morita, M., Kanemori, M., Yanagi, H., and Yura, T. (1999a). Heat-induced synthesis of σ32 in

Escherichia coli: structural and functional dissection of rpoH mRNA secondary structure. J. Bacteriol. *181*, 401–410.

Morita, M.T., Tanaka, Y., Kodama, T.S., Kyogoku, Y., Yanagi, H., and Yura, T. (1999b). Translational induction of heat shock transcription factor σ32: evidence for a built-in RNA thermosensor. Genes Dev. *13*, 655–665.

Narberhaus, F. (2010). Translational control of bacterial heat shock and virulence genes by temperature-sensing mRNAs. RNA Biol. *7*, 84–89.

Narberhaus, F., Käser, R., Nocker, A., and Hennecke, H. (1998). A novel DNA element that controls bacterial heat shock gene expression. Mol. Microbiol. *28*, 315–323.

Neupert, J., and Bock, R. (2009). Designing and using synthetic RNA thermometers for temperature-controlled gene expression in bacteria. Nat. Protoc. *4*, 1262–1273.

Neupert, J., Karcher, D., and Bock, R. (2008). Design of simple synthetic RNA thermometers for temperature-controlled gene expression in Escherichia coli. Nucleic Acids Res. *36*, 1–9.

Nocker, A., Krstulovic, N.P., Perret, X., and Narberhaus, F. (2001). ROSE elements occur in disparate rhizobia and are functionally interchangeable between species. Arch. Microbiol. *176*, 44–51.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rinnenthal, J., Klinkert, B., Narberhaus, F., and Schwalbe, H. (2010). Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. Nucleic Acids Res. *38*, 3834–3847.

Rinnenthal, J., Klinkert, B., Narberhaus, F., and Schwalbe, H. (2011). Modulation of the stability of the Salmonella fourU-type RNA thermometer. Nucleic Acids Res. *39*, 8258–8270.

Shapiro, R.S., and Cowen, L.E. (2012). Thermal control of microbial development and virulence: Molecular mechanisms of microbial temperature sensing. MBio *3*, 1–6.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. *87*, 7668–7672.

Tomoyasu, T., Takaya, A., Sasaki, T., Nagase, T., Kikuno, R., Morioka, M., and Yamamoto, T. (2003). A New Heat Shock Gene, agsA, Which Encodes a Small Chaperone Involved in Suppressing Protein Aggregation in Salmonella enterica Serovar Typhimurium. J. Bacteriol. *185*, 6331–6339.

Uppal, S., Rao Akkipeddi, V.S.N., and Jawali, N. (2008). Posttranscriptional regulation of cspE in Escherichia coli: Involvement of the short 5'-untranslated region. FEMS Microbiol. Lett. *279*, 83–91.

Wagner, D., Rinnenthal, J., Narberhaus, F., and Schwalbe, H. (2015). Mechanistic insights into temperature-dependent regulation of the simple cyanobacterial hsp17 RNA thermometer at base-pair resolution. Nucleic Acids Res. *43*, 5572–5585.

Waldminghaus, T., Fippinger, A., Alfsmann, J., and Narberhaus, F. (2005). RNA thermometers are common in alpha- and gamma-proteobacteria. Biol. Chem. *386*, 1279–1286.

Waldminghaus, T., Gaubig, L.C., and Narberhaus, F. (2007a). Genome-wide bioinformatic prediction and experimental evaluation of potential RNA thermometers. Mol. Genet. Genomics *278*, 555–564.

Waldminghaus, T., Heidrich, N., Brantl, S., and Narberhaus, F. (2007b). FourU: a novel type of RNA thermometer in Salmonella. Mol. Microbiol. *65*, 413–424.

Waldminghaus, T., Kortmann, J., Gesing, S., and Narberhaus, F. (2008). Generation of synthetic RNA-based thermosensors. Biol. Chem. *389*, 1319–1326.

Waldminghaus, T., Gaubig, L.C., Klinkert, B., and Narberhaus, F. (2009). The *Escherichia coli ibpA* thermometer is comprised of stable and unstable structural elements. RNA Biol. *6*, 455–463.

Watters, K.E., Abbott, T.R., and Lucks, J.B. (2015a). Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. *44*.

Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H., and Lucks, J.B. (2015b). Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). METHODS 1–37.

Weber, G.G., Kortmann, J., Narberhaus, F., and Klose, K.E. (2014). RNA thermometer controls temperature-dependent virulence factor expression in Vibrio cholerae. Proc. Natl. Acad. Sci. *111*.

Yamanaka, K., Mitta, M., and Inouye, M. (1999). Mutation analysis of the 5' untranslated region of the cold shock cspA mRNA of Escherichia coli. J. Bacteriol. *181*, 6284–6291.

CHAPTER 6

CONCLUSIONS AND PERSPECTIVES

6.1 Conclusions

In this work I have expanded our understanding of RNA regulators and added to the toolbox of genetic parts available for synthetic biologists. Through my attempts to engineer light-responsive RNAs, I illuminated the capability of SHAPE-Seq to pinpoint locations of ligand-binding in RNA aptamers, and on the way to creating a series of better RNA activators, I developed a suite of generalizable strategies for stabilizing small RNAs. The STARs I helped optimize have filled a gap in the capabilities of RNA genetic regulation, enabling the construction of more sophisticated RNA circuitry, and my creation of theophylline-sensitive aptaSTARs shows the potential to extend STAR function even further. Combined with pre-existing RNA repressors and riboswitches, these regulators could allow for the construction all-RNA networks capable of sensing and responding to multiple small-molecule inputs and performing complex logical manipulations.

In addition, elucidating the mechanism of the naturally occurring *agsA* thermometer has set the stage for the design of future synthetic thermometers and demonstrated the promise of the newly developed in-cell SHAPE-Seq technique. In-cell SHAPE-Seq revealed the subtle differences in *agsA* RNA reactivity at different temperatures, allowing me to probe the structural mechanism of a regulatory RNA in its native context in a way not previously possible. *In vitro* data gave somewhat different results, confirming the importance of studying RNAs in the cellular context and underlining the utility of *in vivo* probing for future research.

6.2 Perspectives on the future of RNA synthetic biology

As a field, synthetic biology holds great promise, but has only recently begun to deliver on its potential—one of the most exciting prospects is the recent report of paper-based diagnostics capable of detecting RNAs associated with Ebola (Pardee et al., 2014) and Zika (Pardee et al., 2016). The diagnostics use paper-mounted freeze-dried cell extract containing a gene network that produces a colorimetric output in response to the presence of a trigger RNA, opening the door for rapid, affordable viral testing. This remarkable advance relies on a few key enabling technologies: RNA toehold switches, accurate nucleic design, and cell-free transcription-translation systems.

The toehold strategy, first reported in DNA by Zhang et al. (2007), relies on the exposure of a small linear region of sequence, the "toehold", designed to seed an interaction with an invading trans-acting trigger sequence that causes large-scale structural rearrangement. Recent work applied this principle to RNA, designing RNA toehold switches where the invading trigger RNA seeded refolding in its target mRNA so as to make the RBS newly accessible, thus activating gene expression (Green et al., 2014) (See Figure 1.4). The design scheme relied on the use of NUPACK, a computational package that can design RNA sequences predicted to fold into particular structures or complexes (Zadeh et al., 2011), easily automating the process of designing new constructs. Combining easily designed RNA toehold switches with a freeze-dried version of the pre-existing cell-free systems for transcription and translation (overviewed in section 1.6) allowed for facile sample addition and increased portability. While work remains to be done in lowering the detection threshold for trigger RNAs and establishing function in clinical samples, paper-based

diagnostics remain an exciting prospect with extremely relevant clinical applications throughout the developing world.

In addition to providing promising healthcare diagnostics, RNA synthetic biology has made recent strides within the realm of metabolic engineering. RBS engineering by Farasat et al. (2014) optimized flow through a carotenoid biosynthesis pathway, while targeted RNA interference (RNAi) (Si et al., 2015) and sRNAs (Na et al., 2013) have been used to increase metabolic flux through a desired pathway by targeting genes in competing pathway for down-regulation. The recently engineered Clustered Regulatory Interspaced Short Palindromic Repeats interference (CRISPRi) method (Qi et al., 2013) allows similar potential for gene knockdown. CRISPRi relies on engineered small guide RNAs (sgRNAs) to direct a catalytically dead variant of the Cas9 enzyme to a complementary DNA target, thereby suppressing transcription. Other techniques extend the potential of specific Cas9 localization by fusing Cas9 to transcription factors (Farzadfard et al., 2013) or fluorescent proteins (Chen et al., 2013) for transcriptional regulation and imaging. There is no doubt that the ability of the engineered Cas9/sgRNA complex to co-localize nearly any protein and nucleic acid will continue to open up future prospects for synthetic biology.

The continued progress of the field will also depend upon the development of more sophisticated modes of gene regulation—timing circuits that generate metabolic pathway enzymes just as their substrates accumulate, decision networks that allow microbes to sense multiple cell markers to determine the delivery of a pharmaceutical payload, and perhaps most importantly, robust modes of control for the containment of the genetic modifications that will enable this type of technology. Ensuring proper controls will

safeguard the future of synthetic biology as a field and enable scientists to address humanity's most pressing needs in sustainability, medicine, and biochemical production.

6.3 Bibliography

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H.M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol. Syst. Biol. *10*, 731.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. ACS Synth. Biol. *2*, 604–613.

Green, A.A., Silver, P.A., Collins, J.J., and Yin, P. (2014). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. Cell *159*, 1–15.

Na, D., Yoo, S.M., Chung, H., Park, H., Park, J.H., and Lee, S.Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. Nat. Biotechnol. *31*, 170–174.

Pardee, K., Green, A.A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., and Collins, J.J. (2014). Paper-Based Synthetic Gene Networks. Cell *159*, 940–954.

Pardee, K., Green, A.A., Takahashi, M.K., Braff, D., Lambert, G., Lee, J.W., Ferrante, T., Ma, D., Donghia, N., Fan, M., et al. (2016). Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell *165*, 1–12.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of

gene expression. Cell *152*, 1173–1183.

Si, T., Luo, Y., Bao, Z., and Zhao, H. (2015). RNAi-assisted genome evolution in saccharomyces cerevisiae for complex phenotype engineering. ACS Synth. Biol. *4*, 283–291.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: analysis and design of nucleic acid systems. J. Comput. Chem. *32*, 170–173.

Zhang, D.Y., Turberfield, A.J., Yurke, B., and Winfree, E. (2007). Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. Science *318*, 1121–1125.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR TOWARD A LIGHT-ACTIVATED RNA REGULATOR (CHAPTER 3)

A.1 Supplementary Figures

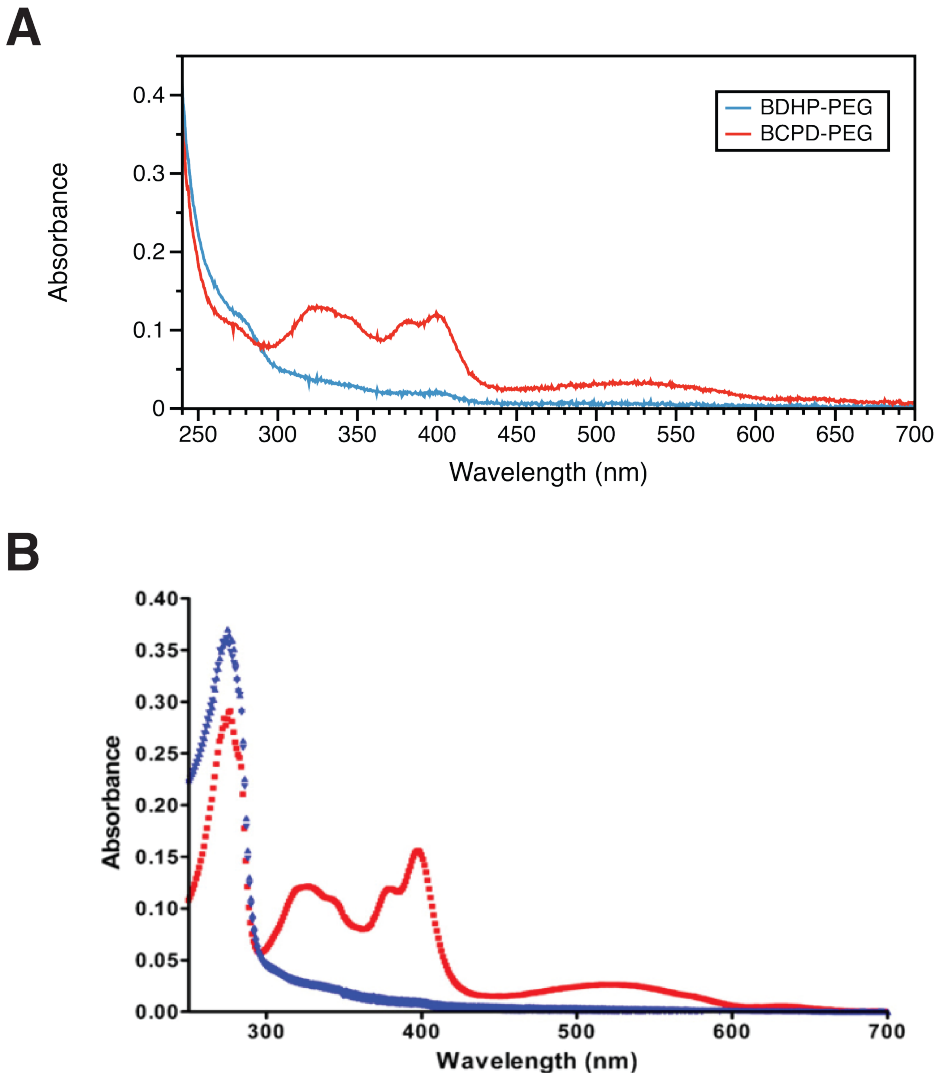


Figure A.1 Absorbance spectra of dihydropyrene ligand. Absorbance spectra of the PEG-attached BDHP and BCDP isomers as **(A)** measured on our samples and **(B)** reported by Lee et al. (2007). *

*Figure reproduced here from original publication.

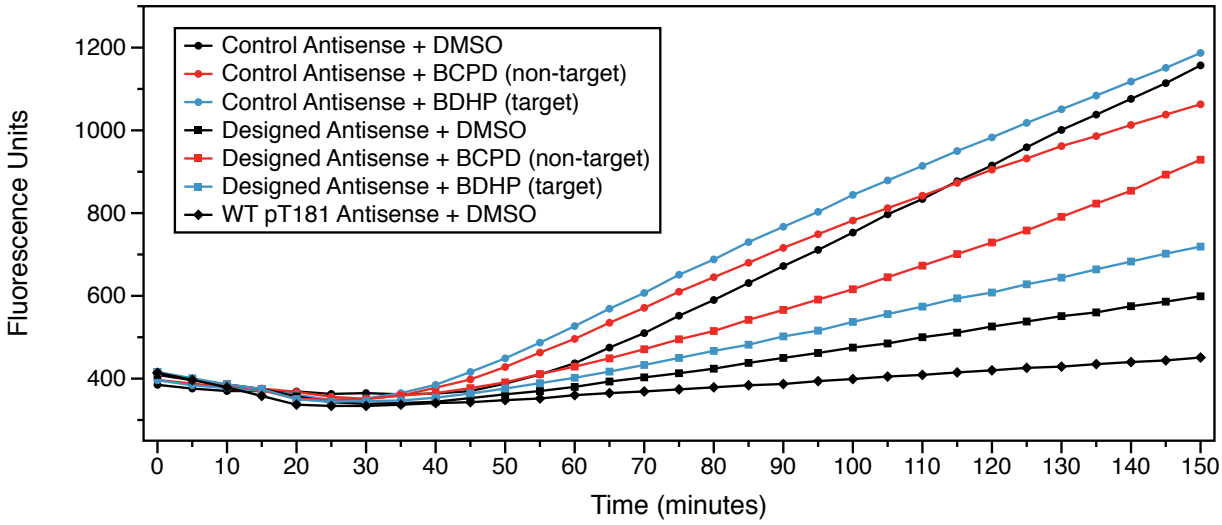


Figure A.2 TX-TL testing of fusion of pT181 antisense and dihydropyrene aptamer. Kinetic fluorescence traces were measured in the TX-TL extract system for various antisense plasmid and ligand combinations, all paired with a pT181 sense plasmid in which the regulator controlled transcription of SFGFP. With the exception of the designed antisense paired with DMSO (ligand-less solvent control), all the traces appeared consistent with BDHP-triggered repression by the designed antisense.

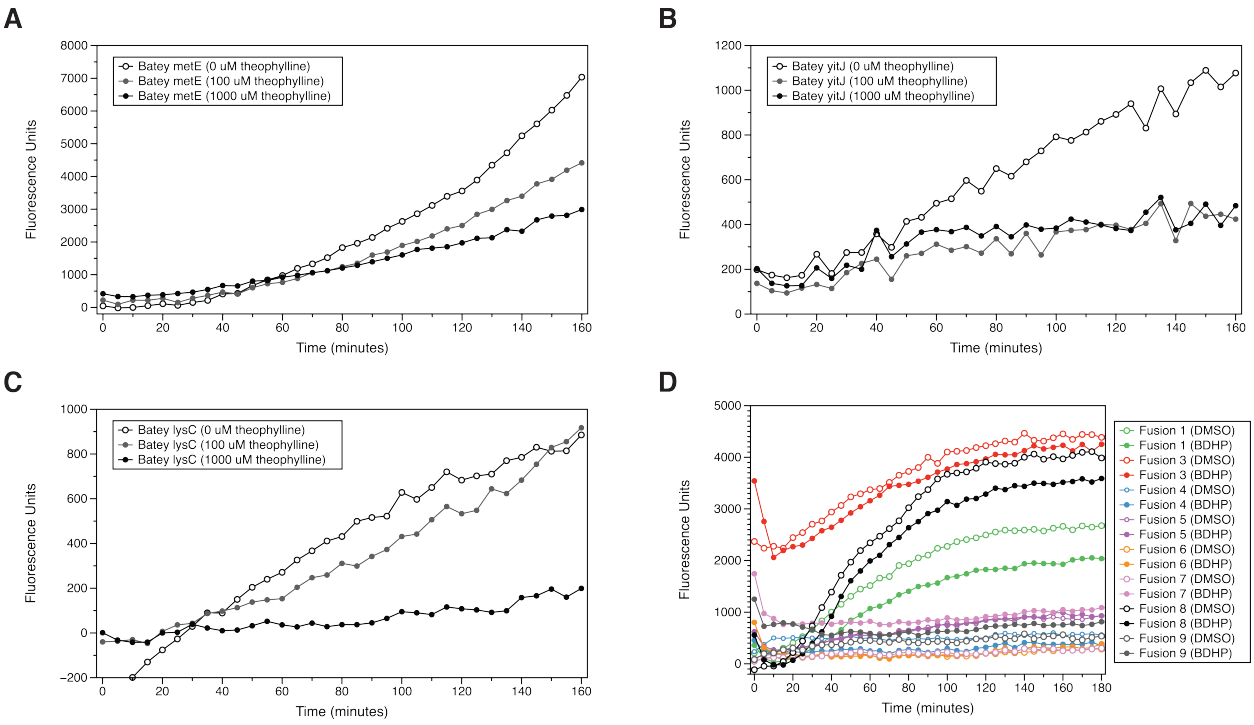


Figure A.3 TX-TL testing of previously reported and newly designed artificial riboswitches. TX-TL confirmation of function of artificial riboswitches using modular expression platforms, previously reported by the Batey Lab (Ceres et al., 2013): **(A)** metE-theophylline riboswitch, **(B)** yitJ-theophylline riboswitch, and **(C)** lysC-theophylline riboswitch. **(D)** Functional testing of various designed fusions to minimal BDHP aptamers using the same riboswitch expression platforms.

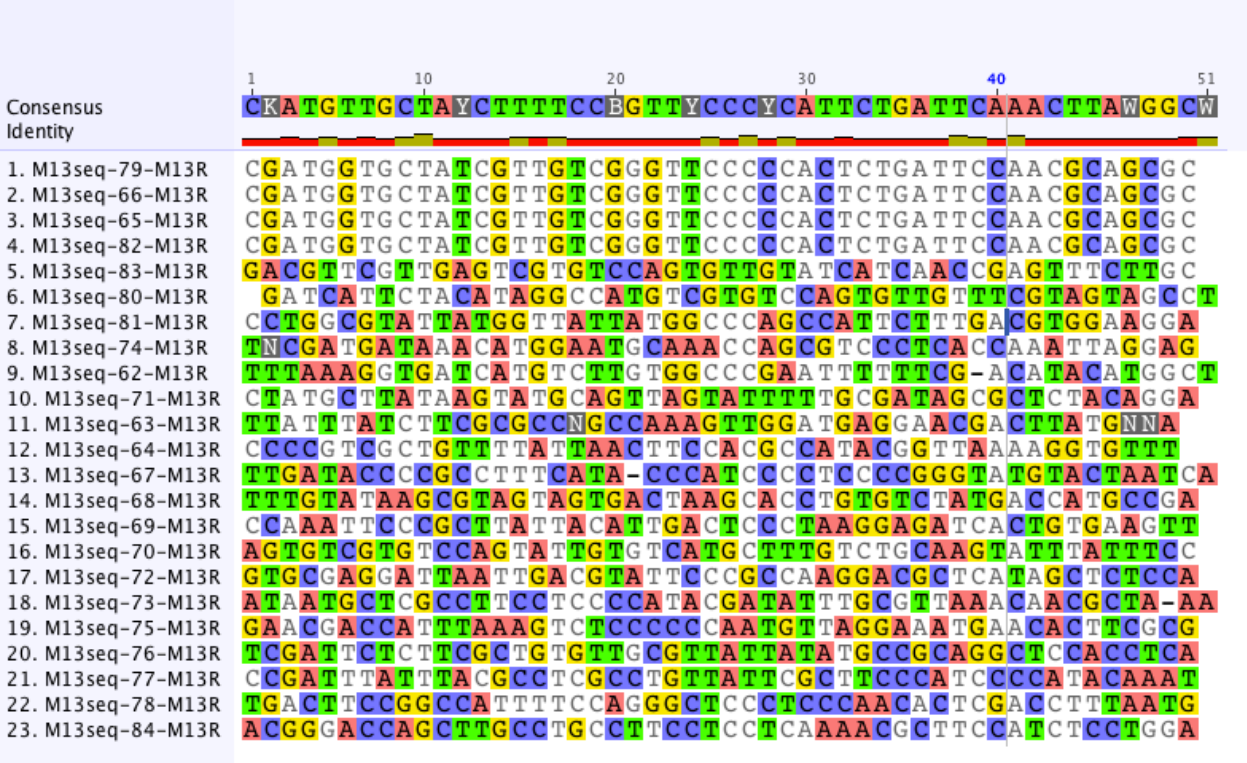


Figure A.4 Sequencing alignment of billiverdin aptamers from round 10 SELEX pool.

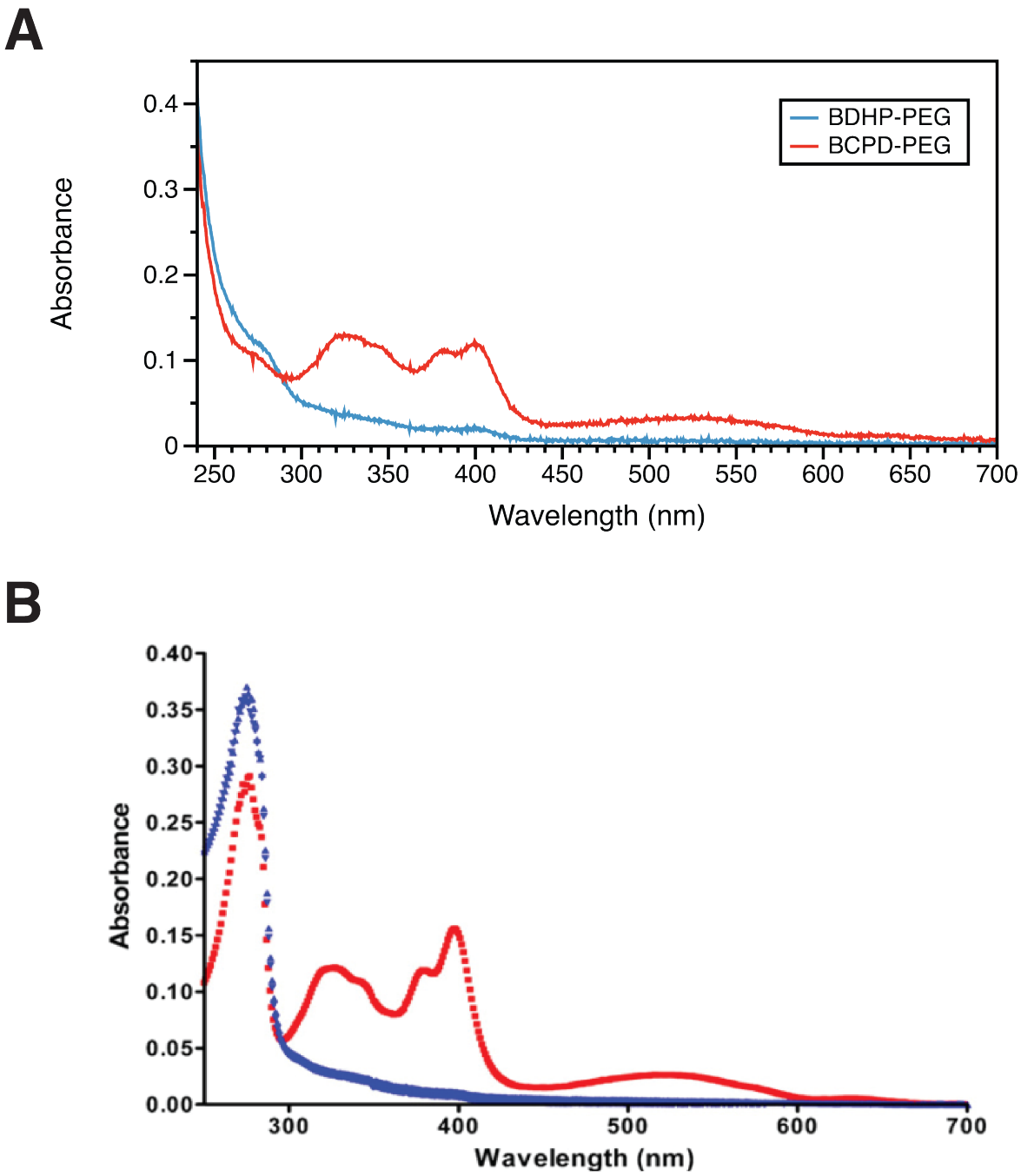


Figure A.5 Absorbance spectra of dihydropyrene ligand. Absorbance spectra of the PEG-attached BDHP and BCDP isomers as **(A)** measured on our samples and **(B)** reported by Lee et al. (2007). *

*Figure reproduced here from original publication.

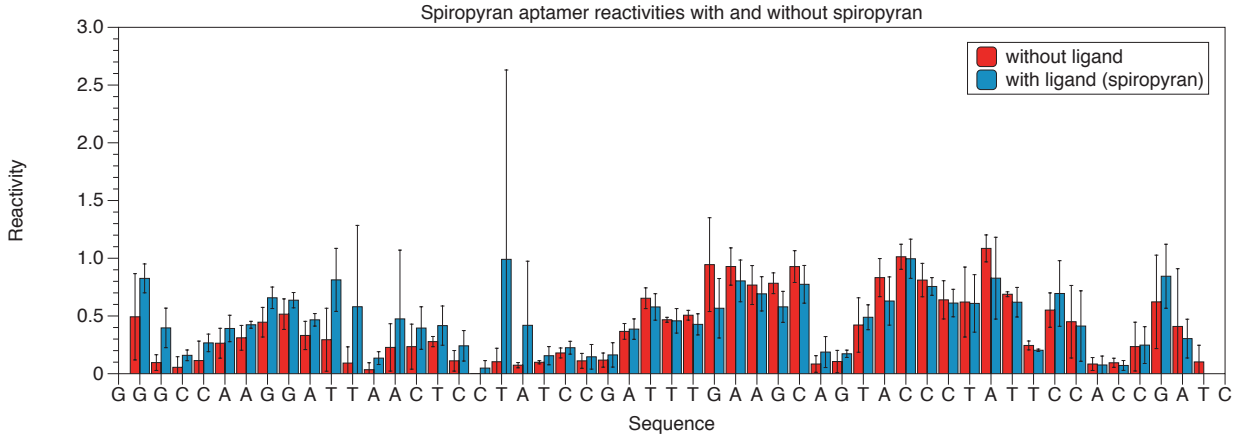


Figure A.6 Spiropyran aptamer SHAPE reactivities with and without ligand. The SHAPE-CE reactivities with and without spiropyran are not statistically different, showing no detectable evidence of ligand binding.

A.2 Supplementary Background: Systematic evolution by exponential enrichment (SELEX)

Systematic evolution by exponential enrichment (SELEX) (Ellington and Szostak, 1990; Tuerk and Gold, 1990) refers to a valuable technique for generating RNA or DNA aptamers specific to a particular ligand. A randomized library of single-stranded DNA or RNAs are incubated with the a filter or resin functionalized with the ligand of interest, and the sequences that do not bind are washed away. Then the binding aptamers are recovered (often through heat elution or competitive binding) for subsequent amplification via PCR (in the case of DNA) or reverse transcription, amplification and transcription (in the case of RNA). The resulting pool of selected aptamers is subjected to additional rounds of the same protocol, generally in the range of 8-20 cycles, until the pool is sufficiently enriched in sequences with high affinity for the target ligand. The pool is then sequenced to determine the particular aptamer sequences that bind the chosen target.

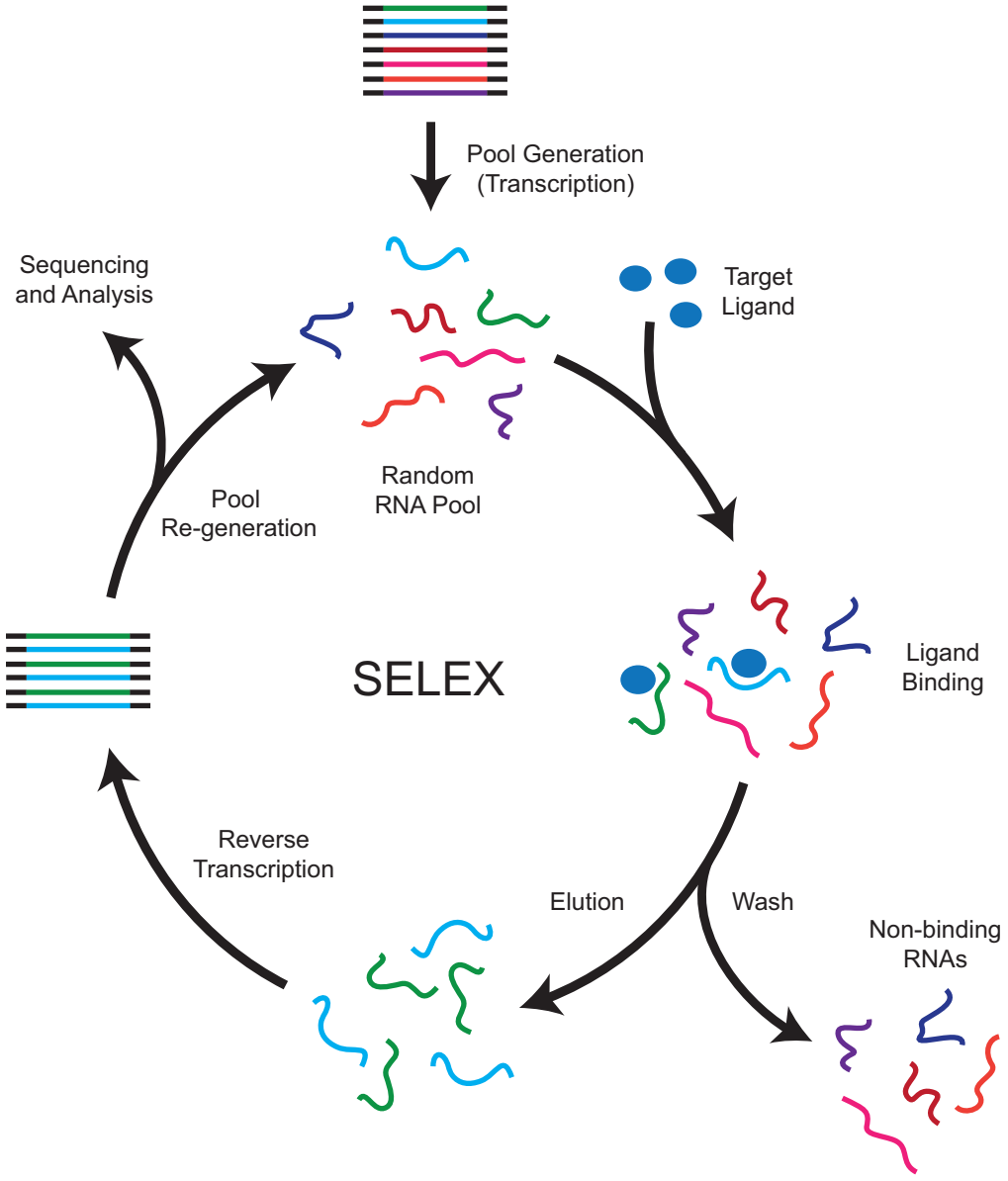


Figure A.7 Systematic evolution by exponential enrichment (SELEX). A random pool of RNAs are transcribed, selected for binding to the target ligand, and reverse transcribed. A selectively enriched RNA pool is then re-generated through PCR amplification and transcription to undergo the selection cycle anew. Successive cycles of SELEX result in a pool of aptamers with strong preference for the target ligand, which can then be sequenced and analyzed.

A.3 Bibliography

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013). Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth. Biol. *2*, 463–472.

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR IMPROVING FOLD ACTIVATION OF SMALL TRANSCRIPTION ACTIVATING RNAS (STARS) WITH RATIONAL RNA ENGINEERING STRATEGIES (CHAPTER 3)

B.1 Supplementary Tables

Table B.1 Plasmids used in this study.

Color code: yellow-promoter, light blue-terminator, dark blue-5' stability hairpin, red-sRNA scaffold, magenta-BamHI-BglII scar, light grey-strong RBS, dark grey-resistance marker, purple-origin of replication, bright green-SFGFP, green-repC fragment (pT181-based systems)

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
JBL001	Control (empty) sense	TrrnB – CmR – p15A origin	GAAGCTTGGGCCCGAACAAAAACTCA TCTCAGAAGAGGATCTGAATAGCGCC GTCGACCATCATCATCATCATTG AGTTTAAACGGTCTCCAGCTTGGCTG TTTTGGCGGATGAGAGAAGATTTTCA GCCTGATACAGATTAAATCAGAACGC AGAAGCGGTCTGATAAAACAGAATTT GCCTGGCGGCAGTAGCGCGGTGGTCC CACCTGACCCCATGCCGAACTCAGAA GTGAAACGCCGTAGCGCCGATGGTAG TGTGGGGTCTCCCCATGCGAGAGTAG GGAACTGCCAGGCATCAAATAAAACG AAAGGCTCAGTCGAAAGACTGGGCCT TTCGTTTTATCTGTTGTTTGTCGGTG AACTAATTATCTAGACTGCAGTTGAT CGGGCACGTAAGAGGTTCCAACTTTC ACCATAATGAAATAAGATCACTACCG GGCGTATTTTTTGAGTTATCGAGATT TTCAGGAGCTAAGGAAGCTAAAATGG AGAAAAAAATCACTGGATATACCACC GTTGATATATCCCAATGGCATCGTAA AGAACATTTTGAGGCATTTCAGTCAG TTGCTCAATGTACCTATAACCAGACC GTTCAGCTGGATATTACGGCCTTTTT AAAGACCGTAAAGAAAAATAAGCAC AAGTTTTATCCGGCCTTTATTCACAT TCTTGCCCGCCTGATGAATGCTCATC CGGAATTTCGTATGGCAATGAAAGAC GGTGAGCTGGTGATATGGGATAGTGT	All (control)

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			TCACCCTTGTTACACCGTTTTCCATG AGCAAACTGAAACGTTTTCATCGCTC TGGAGTGAATACCACGACGATTTCCG GCAGTTTCTACACATATATTCGCAAG ATGTGGCGTGTTACGGTGAAAACCTG GCCTATTTCCCTAAAGGGTTTATTGA GAATATGTTTTTCGTCTCAGCCAATC CCTGGGTGAGTTTCACCAGTTTTGAT TTAAACGTGGCCAATATGGACAACTT CTTCGCCCCCGTTTTCACCATGGGCA AATATTATACGCAAGGCGACAAGGTG CTGATGCCGCTGGCGATTCAGGTTCA TCATGCCGTTTGTGATGGCTTCCATG TCGGCAGAATGCTTAATGAATTACAA CAGTACTGCGATGAGTGGCAGGGCGG GGCGTAATTTGATATCGAGCTCGCTT GGACTCCTGTTGATAGATCCAGTAAT GACCTCAGAACTCCATCTGGATTTGT TCAGAACGCTCGGTTGCCGCCGGGCG TTTTTTATTGGTGAGAATCCAAGCCT CCGATCAACGTCTCATTTTCGCCAAA AGTTGGCCCAGGGCTTCCCGGTATCA ACAGGGACACCAGGATTTATTTATTC TGCGAAGTGATCTTCCGTCACAGGTA TTTATTCGGCGCAAAGTGCGTCGGGT GATGCTGCCAACTTACTGATTTAGTG TATGATGGTGTTTTTGAGGTGCTCCA GTGGCTTCTGTTTCTATCAGCTGTCC CTCCTGTTCAGCTACTGACGGGGTGG TGCGTAACGGCAAAAGCACCGCCGGA CATCAGCGCTAGCGGAGTGTATACTG GCTTACTATGTTGGCACTGATGAGGG TGTCAGTGAAGTGCTTCATGTGGCAG GAGAAAAAAGGCTGCACCGGTGCGTC AGCAGAATATGTGATACAGGATATA TTCCGCTTCCTCGCTCACTGACTCGCT ACGCTCGGTCGTTCGACTGCGGCGAG CGGAAATGGCTTACGAACGGGGCGGA GATTTCCTGGAAGATGCCAGGAAGAT ACTTAACAGGGAAGTGAGAGGGCCGC GGCAAAGCCGTTTTTCCATAGGCTCC GCCCCCCTGACAAGCATCACGAAATC TGACGCTCAAATCAGTGGTGGCGAAA CCCGACAGGACTATAAAGATACCAGG CGTTTCCCCCTGGCGGCTCCCTCGTGC GCTCTCCTGTTCCTGCCTTTCGGTTT ACCGGTGTCATTCCGCTGTTATGGCC	

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			GCGTTTGTCTCATTCCACGCCTGACA CTCAGTTCCGGGTAGGCAGTTCGCTC CAAGCTGGACTGTATGCACGAACCCC CCGTTCAGTCCGACCGCTGCGCCTTA TCCGGTAACTATCGTCTTGAGTCCAA CCCGGAAAGACATGCAAAAGCACCAC TGGCAGCAGCCACTGGTAATTGATTT AGAGGAGTTAGTCTTGAAGTCATGCG CCGGTTAAGGCTAAACTGAAAGGACA AGTTTTGGTGACTGCGCTCCTCCAAG CCAGTTACCTCGGTTCAAAGAGTTGG TAGCTCAGAGAACCTTCGAAAAACCG CCCTGCAAGGCGGTTTTTTCGTTTTC AGAGCAAGAGATTACGCGCAGACCAA AACGATCTCAAGAAGATCATCTTATT AATCAGATAAAATATTTCTAGATTTC AGTGCAATTTATCTCTTCAAATGTAG CACCTGAAGTCAGCCCCATACGATAT AAGTTGTAATTCTCATGTTTGACAGC TTATCATCGATAAGCTTCCGATGGCG CGCCGAGAGGCTTTACACTTTATGCT TCCGGCTGAATTCTAAAGATCCTAAC TCGAGTAATGA	
Sense Plasmid Scaffold	Empty sense plasmid with reporter gene	SFGFP – TrrnB – CmR – p15A origin *All plasmids based on the sense plasmid scaffold backbone have their key sequence inserted directly upstream of SFGFP (bright green) without cloning scars.	ATGAGCAAAGGAGAAGAACTTTTCAC TGGAGTTGTCCCAATTCTTGTTGAAT TAGATGGTGATGTTAATGGGCACAAA TTTTCTGTCCGTGGAGAGGGTGAAGG TGATGCTACAAACGGAAAACTCACCC TTAAATTTATTTGCACTACTGGAAAA CTACCTGTTCCGTGGCCAACACTTGT CACTACTCTGACCTATGGTGTTCAAT GCTTTTCCCGTTATCCGGATCACATG AAACGGCATGACTTTTTCAAGAGTGC CATGCCCGAAGGTTATGTACAGGAAC GCACTATATCTTTCAAAGATGACGGG ACCTACAAGACGCGTGCTGAAGTCAA GTTTGAAGGTGATACCCTTGTTAATC GTATCGAGTTAAAGGGTATTGATTTT AAAGAAGATGGAAACATTCTTGGAC ACAAACTCGAGTACAACTTTAACTCA CACAATGTATACATCACGGCAGACAA ACAAAAGAATGGAATCAAAGCTAACT TCAAAATTCGCCACAACGTTGAAGAT GGTTCCGTTCAACTAGCAGACCATTA TCAACAAAATACTCCAATTGGCGATG GCCCTGTCCTTTTACCAGACAACCAT TACCTGTCGACACAATCTGTCCTTTC	None (for plasmid mapping only)

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			GAAAGATCCCAACGAAAAGCGTGACC ACATGGTCCTTCTTGAGTTTGTAACT GCTGCTGGGATTACACATGGCATGGA TGAGCTCTACAAATAAGGATCTGAAG CTTGGGCCCGAACAAAAACTCATCTC AGAAGAGGATCTGAATAGCGCCGTCG ACCATCATCATCATCATCATTGAGTT TAAACGGTCTCCAGCTTGGCTGTTTT GGCGGATGAGAGAAGATTTTCAGCCT GATACAGATTAAATCAGAACGCAGAA GCGGTCTGATAAAACAGAATTTGCCT GGCGGCAGTAGCGCGGTGGTCCCACC TGACCCCATGCCGAACTCAGAAGTGA AACGCCGTAGCGCCGATGGTAGTGTG GGGTCTCCCCATGCGAGAGTAGGGAA CTGCCAGGCATCAAATAAAACGAAAG GCTCAGTCGAAAGACTGGGCCTTTCG TTTTATCTGTTGTTTGTCGGTGAACT GGATCCTTACTCGAGTCTAGACTGCA GTTGATCGGGCACGTAAGAGGTTCCA ACTTTCACCATAATGAAATAAGATCA CTACCGGGCGTATTTTTTGAGTTATC GAGATTTTCAGGAGCTAAGGAAGCTA AAATGGAGAAAAAAATCACTGGATA TACCACCGTTGATATATCCCAATGGC ATCGTAAAGAACATTTTGAGGCATTT CAGTCAGTTGCTCAATGTACCTATAA CCAGACCGTTCAGCTGGATATTACGG CCTTTTTAAAGACCGTAAAGAAAAAT AAGCACAAGTTTTATCCGGCCTTTAT TCACATTCTTGCCCGCCTGATGAATG CTCATCCGGAATTTCGTATGGCAATG AAAGACGGTGAGCTGGTGATATGGG ATAGTGTTCACCCTTGTTACACCGTT TTCCATGAGCAAACTGAAACGTTTTC ATCGCTCTGGAGTGAATACCACGACG ATTTCCGGCAGTTTCTACACATATAT TCGCAAGATGTGGCGTGTTACGGTGA AAACCTGGCCTATTTCCCTAAAGGGT TTATTGAGAATATGTTTTTCGTCTCA GCCAATCCCTGGGTGAGTTTCACCAG TTTTGATTTAAACGTGGCCAATATGG ACAACTTCTTCGCCCCCGTTTTCACC ATGGGCAAATATTATACGCAAGGCGA CAAGGTGCTGATGCCGCTGGCGATTC AGGTTCATCATGCCGTTTGTGATGGC TTCCATGTCGGCAGAATGCTTAATGA	

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			ATTACAACAGTACTGCGATGAGTGGC AGGGCGGGGCGTAATTTGATATCGAG CTCGCTTGGACTCCTGTTGATAGATC CAGTAATGACCTCAGAACTCCATCTG GATTTGTTCAGAACGCTCGGTTGCCG CCGGGCGTTTTTTATTGGTGAGAATC CAAGCCTCCGATCAACGTCTCATTTT CGCCAAAAGTTGGCCCAGGGCTTCCC GGTATCAACAGGGACACCAGGATTTA TTTATTCTGCGAAGTGATCTTCCGTC ACAGGTATTTATTCGGCGCAAAGTGC GTCGGGTGATGCTGCCAACTTACTGA TTTAGTGTATGATGGTGTTTTTGAGG TGCTCCAGTGGCTTCTGTTTCTATCA GCTGTCCCTCCTGTTCAGCTACTGAC GGGGTGGTGCGTAACGGCAAAAGCAC CGCCGGACATCAGCGCTAGCGGAGTG TATACTGGCTTACTATGTTGGCACTG ATGAGGGTGTCAGTGAAGTGCTTCAT GTGGCAGGAGAAAAAAGGCTGCACCG GTGCGTCAGCAGAATATGTGATACAG GATATATTCCGCTTCCTCGCTCACTG ACTCGCTACGCTCGGTCGTTCGACTG CGGCGAGCGGAAATGGCTTACGAACG GGGCGGAGATTTCCTGGAAGATGCCA GGAAGATACTTAACAGGGAAGTGAG AGGGCCGCGGCAAAGCCGTTTTTCCA TAGGCTCCGCCCCCCTGACAAGCATC ACGAAATCTGACGCTCAAATCAGTGG TGGCGAAACCCGACAGGACTATAAAG ATACCAGGCGTTTCCCCCTGGCGGCT CCCTCGTGCGCTCTCCTGTTCCTGCCT TTCGGTTTACCGGTGTCATTCCGCTG TTATGGCCGCGTTTGTCTCATTCCAC GCCTGACACTCAGTTCCGGGTAGGCA GTTCGCTCCAAGCTGGACTGTATGCA CGAACCCCCCGTTCAGTCCGACCGCT GCGCCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGGAAAGACATGCAAA AGCACCACTGGCAGCAGCCACTGGTA ATTGATTTAGAGGAGTTAGTCTTGAA GTCATGCGCCGGTTAAGGCTAAACTG AAAGGACAAGTTTTGGTGACTGCGCT CCTCCAAGCCAGTTACCTCGGTTCAA AGAGTTGGTAGCTCAGAGAACCTTCG AAAAACCGCCCTGCAAGGCGGTTTTT TCGTTTTCAGAGCAAGAGATTACGCG	

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			CAGACCAAAACGATCTCAAGAAGATC ATCTTATTAATCAGATAAAATATTTC TAGATTTCAGTGCAATTTATCTCTTC AAATGTAGCACCTGAAGTCAGCCCCA TACGATATAAGTTGTAATTCTCATGT TTGACAGCTTATCATCGATAAGCTTC CGATGGCGCGCCGAGAGGCTTTACAC TTTATGCTTCCGGCTGAATTCTAAAG ATCC	
JBL002	Control (empty) antisense	J23119 SpeI – TrrnB – ColE1 origin – AmpR *All plasmids based on the JBL002 backbone have their key sequence inserted in place of J23119 SpeI and TrrnB.	TTGACAGCTAGCTCAGTCCTAGGTAT AATACTAGTGGATCTGAAGCTTGGGC CCGAACAAAAACTCATCTCAGAAGAG GATCTGAATAGCGCCGTCGACCATCA TCATCATCATCATTGAGTTTAAACGG TCTCCAGCTTGGCTGTTTTGGCGGAT GAGAGAAGATTTTCAGCCTGATACAG ATTAAATCAGAACGCAGAAGCGGTCT GATAAAACAGAATTTGCCTGGCGGCA GTAGCGCGGTGGTCCCACCTGACCCC ATGCCGAACTCAGAAGTGAAACGCCG TAGCGCCGATGGTAGTGTGGGGTCTC CCCATGCGAGAGTAGGGAACTGCCAG GCATCAAATAAAACGAAAGGCTCAGT CGAAAGACTGGGCCTTTCGTTTTATC TGTTGTTTGTCGGTGAACTGGATCCT TACTCGAGTCTAGACTGCAGGCTTCC TCGCTCACTGACTCGCTGCGCTCGGT CGTTCGGCTGCGGCGAGCGGTATCAG CTCACTCAAAGGCGGTAATACGGTTA TCCACAGAATCAGGGGATAACGCAGG AAAGAACATGTGAGCAAAAGGCCAGC AAAAGGCCAGGAACCGTAAAAAGGCC GCGTTGCTGGCGTTTTTCCACAGGCT CCGCCCCCCTGACGAGCATCACAAAA ATCGACGCTCAAGTCAGAGGTGGCGA AACCCGACAGGACTATAAAGATACCA GGCGTTTCCCCCTGGAAGCTCCCTCG TGCGCTCTCCTGTTCCGACCCTGCCGC TTACCGGATACCTGTCCGCCTTTCTC CCTTCGGGAAGCGTGGCGCTTTCTCA TAGCTCACGCTGTAGGTATCTCAGTT CGGTGTAGGTCGTTCGCTCCAAGCTG GGCTGTGTGCACGAACCCCCCGTTCA GCCCGACCGCTGCGCCTTATCCGGTA ACTATCGTCTTGAGTCCAACCCGGTA AGACACGACTTATCGCCACTGGCAGC AGCCACTGGTAACAGGATTAGCAGAG	All (subtracted growth control and no antisense condition)

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			CGAGGTATGTAGGCGGTGCTACAGAG TTCTTGAAGTGGTGGCCTAACTACGG CTACACTAGAAGAACAGTATTTGGTA TCTGCGCTCTGCTGAAGCCAGTTACC TTCGGAAAAAGAGTTGGTAGCTCTTG ATCCGGCAAACAAACCACCGCTGGTA GCGGTGGTTTTTTTGTTTGCAAGCAG CAGATTACGCGCAGAAAAAAAGGATC TCAAGAAGATCCTTTGATCTTTTCTA CGGGGTCTGACGCTCAGTGGAACGAA AACTCACGTTAAGGGATTTTGGTCAT GAGATTATCAAAAAGGATCTTCACCT AGATCCTTTTAAATTAAAAATGAAGT TTTAAATCAATCTAAAGTATATATGA GTAAACTTGGTCTGACAGTTACCAAT GCTTAATCAGTGAGGCACCTATCTCA GCGATCTGTCTATTTCGTTCATCCAT AGTTGCCTGACTCCCCGTCGTGTAGA TAACTACGATACGGGAGGGCTTACCA TCTGGCCCCAGTGCTGCAATGATACC GCGAGACCCACGCTCACCGGCTCCAG ATTTATCAGCAATAAACCAGCCAGCC GGAAGGGCCGAGCGCAGAAGTGGTCC TGCAACTTTATCCGCCTCCATCCAGT CTATTAATTGTTGCCGGGAAGCTAGA GTAAGTAGTTCGCCAGTTAATAGTTT GCGCAACGTTGTTGCCATTGCTACAG GCATCGTGGTGTCACGCTCGTCGTTT GGTATGGCTTCATTCAGCTCCGGTTC CCAACGATCAAGGCGAGTTACATGAT CCCCCATGTTGTGCAAAAAAGCGGTT AGCTCCTTCGGTCCTCCGATCGTTGT CAGAAGTAAGTTGGCCGCAGTGTTAT CACTCATGGTTATGGCAGCACTGCAT AATTCTCTTACTGTCATGCCATCCGT AAGATGCTTTTCTGTGACTGGTGAGT ACTCAACCAAGTCATTCTGAGAATAG TGTATGCGGCGACCGAGTTGCTCTTG CCCGGCGTCAATACGGGATAATACCG CGCCACATAGCAGAACTTTAAAAGTG CTCATCATTGGAAAACGTTCTTCGGG GCGAAAACTCTCAAGGATCTTACCGC TGTTGAGATCCAGTTCGATGTAACCC ACTCGTGCACCCAACTGATCTTCAGC ATCTTTTACTTTCACCAGCGTTTCTG GGTGAGCAAAAACAGGAAGGCAAAA TGCCGCAAAAAAGGGAATAAGGGCGA	

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			CACGGAAATGTTGAATACTCATACTC TTCCTTTTTCAATATTATTGAAGCAT TTATCAGGGTTATTGTCTCATGAGCG GATACATATTTGAATGTATTTAGAAA AATAAACAAATAGGGGTTCCGCGCAC ATTTCCCCGAAAAGTGCCACCTGACG TCTAAGAAACCATTATTATCATGACA TTAACCTATAAAAATAGGCGTATCAC GAGGCAGAATTTCAGATAAAAAAAA TCCTTAGCTTTCGCTAAGGATGATTT CTGGAATTCTAAAGATCT	
JBL2184	WT pbuE sense (Chappell et al., 2015)	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtG TGTCTACCAGGAACCGTAAAATCCTG ATTACAAAATTTGTTTATGACATTTT TTGTAATCAGGATTTTTTTTggatctag gaggaaggatct	3.1B, 3.2B, 3.3B-E, 3.4A, B.2
JBL2191	WT pbuE antisense (Chappell et al., 2015)	JBL002	ttgacagctagctcagtcctaggtataatactagtA TAAACAAATTTTGTAATCAGGATTTT ACGGTTCCTGGTAGACACggatctcaaa GCCCGCCgaaaGGCGGGCttttttt	3.1B, 3.2B, 3.3B-E, B.2, B.3
JBL1737	pbuE sense with J23100 promoter	Sense plasmid backbone	ttgacggctagctcagtcctaggtacagtgctagc GTGTCTACCAGGAACCGTAAAATCCT GATTACAAAATTTGTTTATGACATTT TTTGTAATCAGGATTTTTTTTggatcta ggaggaaggatct	3.2B, 3.4A
JBL1738	pbuE sense with J23102 promoter	Sense plasmid backbone	ttgacagctagctcagtcctaggtactgtgctagc GTGTCTACCAGGAACCGTAAAATCCT GATTACAAAATTTGTTTATGACATTT TTTGTAATCAGGATTTTTTTTggatcta ggaggaaggatct	3.2B, 3.4A
JBL1739	pbuE sense with J23119 promoter	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatgctagc GTGTCTACCAGGAACCGTAAAATCCT GATTACAAAATTTGTTTATGACATTT TTTGTAATCAGGATTTTTTTTggatcta ggaggaaggatct	3.2B, 3.4A
JBL1740	pbuE sense with J23150 promoter	Sense plasmid backbone	tttacggctagctcagtcctaggtattatgctagcG TGTCTACCAGGAACCGTAAAATCCTG ATTACAAAATTTGTTTATGACATTTT TTGTAATCAGGATTTTTTTTggatctag gaggaaggatct	3.2B, 3.4A
JBL1741	pbuE sense with J23151 promoter	Sense plasmid backbone	ttgatggctagctcagtcctaggtacaatgctagc GTGTCTACCAGGAACCGTAAAATCCT GATTACAAAATTTGTTTATGACATTT TTTGTAATCAGGATTTTTTTTggatcta ggaggaaggatct	3.2B, 3.4A, 3.5, B.7
JBL3087	pbuE	JBL002	ttgacggctagctcagtcctaggtacagtgctagc	B.2

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	antisense with J23100 promoter		ATAAACAAATTTTGTAATCAGGATTT TACGGTTCCTGGTAGACACggatctcaa aGCCCGCCgaaaGGCGGGCttttttt	
JBL3088	pbuE antisense with J23102 promoter	JBL002	ttgacagctagctcagtcctaggtactgtgctagc ATAAACAAATTTTGTAATCAGGATTT TACGGTTCCTGGTAGACACggatctcaa aGCCCGCCgaaaGGCGGGCttttttt	B.2
JBL3089	pbuE antisense with J23119 promoter	JBL002	ttgacagctagctcagtcctaggtataatgctagc ATAAACAAATTTTGTAATCAGGATTT TACGGTTCCTGGTAGACACggatctcaa aGCCCGCCgaaaGGCGGGCttttttt	B.2
JBL3090	pbuE antisense with J23150 promoter	JBL002	tttacggctagctcagtcctaggtattatgctagcA TAAACAAATTTTGTAATCAGGATTTT ACGGTTCCTGGTAGACACggatctcaaa GCCCGCCgaaaGGCGGGCttttttt	B.2
JBL3091	pbuE antisense with J23151 promoter	JBL002	ttgatggctagctcagtcctaggtacaatgctagc ATAAACAAATTTTGTAATCAGGATTT TACGGTTCCTGGTAGACACggatctcaa aGCCCGCCgaaaGGCGGGCttttttt	B.2
JBL1765	pbuE antisense with 5' pHP14 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgactctcgagtgagattgttgacggtaccgta ttttATAAACAAATTTTGTAATCAGGA TTTTACGGTTCCTGGTAGACACggatct caaaGCCCGCCgaaaGGCGGGCttttttt	3.3B, B.3
JBL1766	pbuE antisense with 5' pHP16 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagtctctgcagagattcgatg atattgttgacggtaccgtatttttATAAACAAA TTTTGTAATCAGGATTTTACGGTTCC TGGTAGACACggatctcaaaGCCCGCCga aaGGCGGGCttttttt	3.3B, B.3
JBL1767	pbuE antisense with 5' pHP17 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagactgcagttcaatagagat attgttgacggtaccgtatttttATAAACAAAT TTTGTAATCAGGATTTTACGGTTCCT GGTAGACACggatctcaaaGCCCGCCgaa aGGCGGGCttttttt	3.3B
JBL1775	pbuE antisense with 3' MicF sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagtA TAAACAAATTTTGTAATCAGGATTTT ACGGTTCCTGGTAGACACggatctCGTC ATTCATTTCTGAATGTCTGTTTACCC CTATTTCAACCGGATGCCTCGCATTC	3.3C, E

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
			GGTTTTTTTTTcaaaGCCCGCCgaaaGGC GGGCtttttttt	
JBL1777	pbuE antisense with 3' Dsra1.3 sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagtA TAAACAAATTTTGTAATCAGGATTTT ACGGTTCCTGGTAGACACggatctGAAT TTTTTAAGTGCccCcgGCTTcgGCggGc TTCATCCCGgCCCCCTCAGGGcCGGGA TTTTTTTTcaaaGCCCGCCgaaaGGCGGG Ctttttttt	3.3C, E
JBL1778	pbuE antisense with 3' Spot42 sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagtA TAAACAAATTTTGTAATCAGGATTTT ACGGTTCCTGGTAGACACggatctGATT TGGCTGAATATTTTAGCCGCCCCAGT CAGTAATGACTGGGGCGTTTTTTcaaa GCCCGCCgaaaGGCGGGCttttttt	3.3C-E, B3
JBL1668	pbuE antisense with 3' MicF sRNA scaffold and 5' pHP14 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgactctcgagtgagattgttgacggtaccgta tttttATAAACAAATTTTGTAATCAGGA TTTTACGGTTCCTGGTAGACACggatct CGTCATTCATTTCTGAATGTCTGTTT ACCCCTATTTCAACCGGATGCCTCGC ATTCGGTTTTTTTTcaaaGCCCGCCgaa aGGCGGGCttttttt	3.3E
JBL1681	pbuE antisense with 3' MicF sRNA scaffold and 5' pHP16 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagtctctgcagagattcgatg atattgttgacggtaccgtatttttATAAACAAA TTTTGTAATCAGGATTTTACGGTTCC TGGTAGACACggatctCGTCATTCATTT CTGAATGTCTGTTTACCCCTATTTCA ACCGGATGCCTCGCATTCGGTTTTTT TTcaaaGCCCGCCgaaaGGCGGGCttttttt t	3.3E
JBL1684	pbuE antisense with 3' MicF sRNA scaffold and 5' pHP17 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagactgcagttcaatagagat attgttgacggtaccgtatttttATAAACAAAT TTTGTAATCAGGATTTTACGGTTCCT GGTAGACACggatctCGTCATTCATTTC TGAATGTCTGTTTACCCCTATTTCAA CCGGATGCCTCGCATTCGGTTTTTTT TcaaaGCCCGCCgaaaGGCGGGCttttttt	3.3E
JBL1725	pbuE antisense with 3' Dsra1.3 sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgactctcgagtgagattgttgacggtaccgta tttttATAAACAAATTTTGTAATCAGGA TTTTACGGTTCCTGGTAGACACggatct GAATTTTTTAAGTGCccCcgGCTTcgG CggGcTTCATCCCGgCCCCCTCAGGGcC	3.3E

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	and 5' pHP14 hairpin		GGGATTTTTTTcaaaGCCCGCCgaaaGGCGGGCtttttttt	
JBL1726	pbuE antisense with 3' Dsra1.3 sRNA scaffold and 5' pHP16 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgacttatctcgagtctctgcagagattcgatgatattgttgacggtaccgtatttttATAAACAAATTTTTGTAATCAGGATTTTACGGTTCCTGGTAGACACggatctGAATTTTTTTAAGTGCccCcgGCTTcgGCggGcTTCATCCCGgCCCCCTCAGGGcCGGGATTTTTTTTcaaaGCCCGCCgaaaGGCGGGCtttttttt	3.3E
JBL1792	pbuE antisense with 3' Dsra1.3 sRNA scaffold and 5' pHP17 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgacttatctcgagactgcagttcaatagagatattgttgacggtaccgtatttttATAAACAAATTTTTGTAATCAGGATTTTACGGTTCCTGGTAGACACggatctGAATTTTTTAAGTGCccCcgGCTTcgGCggGcTTCATCCCGgCCCCCTCAGGGcCGGGATTTTTTTTcaaaGCCCGCCgaaaGGCGGGCtttttttt	3.3E
JBL1793	pbuE antisense with 3' Spot42 sRNA scaffold and 5' pHP14 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgactctcgagtgagattgttgacggtaccgtatttttATAAACAAATTTTGTAATCAGGATTTTACGGTTCCTGGTAGACACggatctGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGTAATGACTGGGGCGTTTTTTTcaaaGCCCGCCgaaaGGCGGGCtttttttt	3.3E
JBL1794	pbuE antisense with 3' Spot42 sRNA scaffold and 5' pHP16 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgacttatctcgagtctctgcagagattcgatgatattgttgacggtaccgtatttttATAAACAAATTTTGTAATCAGGATTTTACGGTTCCTGGTAGACACggatctGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGTAATGACTGGGGCGTTTTTTTcaaaGCCCGCCgaaaGGCGGGCtttttttt	3.3E, 3.4A, 3.5, B.7
JBL1795	pbuE antisense with 3' Spot42 sRNA scaffold and 5'	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgacttatctcgagactgcagttcaatagagatattgttgacggtaccgtatttttATAAACAAATTTTGTAATCAGGATTTTACGGTTCCTGGTAGACACAgatctGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGTAATGACTGGGGCGTTTTTTTcaaaGCCCGCCg	3.3E

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	pHP17 hairpin		aaaGGCGGGCttttttttt	
JBL2111	WT prgX sense (Chappell et al., 2015)	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtA ATGTTGAGCAGCGGGGAATGTATACA GTTCATGTATATGTTCCCCGCTTTTT TTTggatctaggaggaaggatct	3.4B, B.6
JBL2117	WT prgX antisense (Chappell et al., 2015)	JBL002	ttgacagctagctcagtcctaggtataatactagtT GAACTGTATACATTCCCCGCTGCTCA ACATTggatctcaaaGCCCGCCgaaaGGC GGGCttttttttt	3.4B, B.6
JBL3010	prgX antisense with 5' pHP14 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgactctcgagtgagattgttgacggtaccgta tttttGAACTGTATACATTCCCCGCTGC TCAACATTggatctcaaaGCCCGCCgaaa GGCGGGCttttttttt	3.4B, 3.4C, 3.5, B.7
JBL3011	prgX antisense with 5' pHP16 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagtctctgcagagattcgatg atattgttgacggtaccgtatttttGAACTGTA TACATTCCCCGCTGCTCAACATTggat ctcaaaGCCCGCCgaaaGGCGGGCttttttttt	3.4B
JBL3012	prgX antisense with 5' pHP17 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagactgcagttcaatagagat attgttgacggtaccgtatttttGAACTGTATA CATTCCCCGCTGCTCAACATTggatctc aaaGCCCGCCgaaaGGCGGGCttttttttt	3.4B
JBL3022	prgX antisense with 3' Spot42 sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagtT GAACTGTATACATTCCCCGCTGCTCA ACATTggatctGATTTGGCTGAATATT TTAGCCGCCCCAGTCAGTAATGACTG GGGCGTTTTTTcaaaGCCCGCCgaaaGG CGGGCttttttttt	B.6
JBL3026	prgX antisense with 3' Spot42 sRNA scaffold and 5' pHP14 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgactctcgagtgagattgttgacggtaccgta tttttGAACTGTATACATTCCCCGCTGC TCAACATTggatctGATTTGGCTGAAT ATTTTAGCCGCCCCAGTCAGTAATGA CTGGGGCGTTTTTTcaaaGCCCGCCgaa aGGCGGGCttttttttt	B.6
JBL3027	prgX antisense with 3' Spot42 sRNA scaffold	JBL002	ttgacagctagctcagtcctaggtataatactagta cgtcgacttatctcgagtctctgcagagattcgatg atattgttgacggtaccgtatttttGAACTGTA TACATTCCCCGCTGCTCAACATTggat ctGATTTGGCTGAATATTTTAGCCGCC CCAGTCAGTAATGACTGGGGCGTTTT	B.6

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	and 5' pHP16 hairpin		TTcaaaGCCCGCCgaaaGGCGGGCttttttt	
JBL3028	prgX antisense with 3' Spot42 sRNA scaffold and 5' pHP17 hairpin	JBL002	ttgacagctagctcagtcctaggtataatactagtacgtcgacttatctcgagactgcagttcaatagagatattgttgacggtaccgtatttTGAACTGTATACATTCCCCGCTGCTCAACATTggatctGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGTAATGACTGGGGCGTTTTTTcaaaGCCCGCCgaaaGGCGGGCttttttt	B.6
JBL3038	prgX sense with J23100 promoter	Sense plasmid backbone	ttgacggctagctcagtcctaggtacagtgctagcAATGTTGAGCAGCGGGGAATGTATACAGTTCATGTATATGTTCCCCGCTTTTTTTTggatctaggaggaaggatct	3.4C
JBL3039	prgX sense with J23102 promoter	Sense plasmid backbone	ttgacagctagctcagtcctaggtactgtgctagcAATGTTGAGCAGCGGGGAATGTATACAGTTCATGTATATGTTCCCCGCTTTTTTTTggatctaggaggaaggatct	3.4C
JBL3040	prgX sense with J23119 promoter	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatgctagcAATGTTGAGCAGCGGGGAATGTATACAGTTCATGTATATGTTCCCCGCTTTTTTTTggatctaggaggaaggatct	3.4C
JBL3041	prgX sense with J23150 promoter	Sense plasmid backbone	tttacggctagctcagtcctaggtattatgctagcAATGTTGAGCAGCGGGGAATGTATACAGTTCATGTATATGTTCCCCGCTTTTTTTTggatctaggaggaaggatct	3.4C
JBL3042	prgX sense with J23151 promoter	Sense plasmid backbone	ttgatggctagctcagtcctaggtacaatgctagcAATGTTGAGCAGCGGGGAATGTATACAGTTCATGTATATGTTCCCCGCTTTTTTTTggatctaggaggaaggatct	3.4C, 3.5, B.7
JBL007	pT181.H1 attenuator sense (Lucks et al., 2011)	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtAACAAAATAAAAAGGAGTCGCTCTGTCCCTCGCCAAAGTTGCAGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTTTTTATAATCTTGTATATTTAGATATTAAACGATATTTAAATATACATAAAGATATATATTTGGGTGAGcgattccttaaacgaaattgagattaaggagtcgCtCttttttatgtataaaaacaatcatgcaaatcattcaaatcatttggaaaatcacgatttagacaatttttctaaaaccggctactctaatagccggttgtaaggatctaggaggaaggatct	3.5, B.7
JBL008	pT181.H1 attenuator	JBL002	ttgacagctagctcagtcctaggtataatactagtATACAAGATTATAAAAACAACTCAGTG	3.5, B.7

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	antisense (Lucks et al., 2011)		TTTTTTTCTTTGAATGATGTCGTTCT GCAACTTTGGCGAGGGACAGAGCGAC TCCTTTTTATTTggatctgaagcttgggccc gaacaaaaactcatctcagaagaggatctgaata gcgccgtcgaccatcatcatcatcattgagttt aaacggtctccagcttggctgttttggcggatgag agaagattttcagcctgatacagattaaatcagaa cgcagaagcggtctgataaaacagaatttgcctgg cggcagtagcgcggtggtcccacctgaccccatgc cgaactcagaagtgaaacgccgtagcgccgatgg tagtgtggggtctccccatgcgagagtagggaact gccaggcatcaaataaaacgaaaggctcagtcga aagactgggcctttcgttttatctgttgtttgtcggtg aact	
JBL1020	Fusion 6 attenuator sense (Takahashi and Lucks, 2013)	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtA ACAAAATAAAAAGGAGTCGCTCACTT ACGAACTTGGCGGAACGACGTGAACG ACATCATTCAAAGAAAAAAACACTGA GTTGTTTTTATAATCTTGTATATTTA GATATTAAACGATATTTAAATATACA TAAAGATATATATTTGGGTGAGcgatt ccttaaacgaaattgagattaaggagtcgCtCtttt ttatgtataaaaacaatcatgcaaatcattcaaatc atttggaaaatcacgatttagacaatttttctaaaa ccggctactctaatagccggttgtaaggatctagg aggaaggatct	3.5, B.7
JBL1029	Fusion 6 attenuator antisense (Takahashi and Lucks, 2013)	JBL002	ttgacagctagctcagtcctaggtataatactagtA TACAAGATTATAAAAACAACTCAGTG TTTTTTTCTTTGAATGATGTCGTTCA CGTCGTTCCGCCAAGTTCGTAAGTGA GCGACTCCTTTTTATTTggatctgaagctt gggcccgaacaaaaactcatctcagaagaggatc tgaatagcgccgtcgaccatcatcatcatcatcatt gagtttaaacggtctccagcttggctgttttggcgg atgagagaagattttcagcctgatacagattaaat cagaacgcagaagcggtctgataaaacagaattt gcctggcggcagtagcgcggtggtcccacctgac cccatgccgaactcagaagtgaaacgccgtagcg ccgatggtagtgtggggtctccccatgcgagagta gggaactgccaggcatcaaataaaacgaaaggct cagtcgaaagactgggcctttcgttttatctgttgtt tgtcggtgaact	3.5, B.7
JBL1126	Fusion 4 attenuator sense (Takahashi and Lucks,	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtA ACAAAATAAAAAGGAGTCGCTCACGT TCAACTTTGGCGAGTACGATGTGAAC GACATCATTCAAAGAAAAAAACACTG AGTTGTTTTTATAATCTTGTATATTT	3.5, B.7

Plasmid Number	Name/ Reference	Plasmid backbone/ architecture	Key sequence	Figures
	2013)		AGATATTAAACGATATTTAAATATAC ATAAAGATATATATTTGGGTGAGcgat tccttaaacgaaattgagattaaggagtcgCtCttt tttatgtataaaaacaatcatgcaaatcattcaaat catttggaaaatcacgatttagacaatttttctaaa accggctactctaatagccggttgtaaggatctag gaggaaggatct	
JBL1033	Fusion 4 attenuator antisense (Takahashi and Lucks, 2013)	JBL002	ttgacagctagctcagtcctaggtataatactagtA TACAAGATTATAAAAACAACTCAGTG TTTTTTTCTTTGAATGATGTCGTTCA CATCGTACTCGCCAAAGTTGAACGTG AGCGACTCCTTTTTTATTTggatctgaagc ttgggcccgaacaaaaactcatctcagaagagga tctgaatagcgccgtcgaccatcatcatcatcatca ttgagtttaaacggtctccagcttggctgttttggcg gatgagagaagattttcagcctgatacagattaaa tcagaacgcagaagcggtctgataaaacagaatt tgcctggcggcagtagcgcggtggtcccacctgac cccatgccgaactcagaagtgaaacgccgtagcg ccgatggtagtgtggggtctccccatgcgagagta gggaactgccaggcatcaaataaaacgaaaggct cagtcgaaagactgggcctttcgttttatctgttgtt tgtcggtgaact	3.5, B.7
JBL2844	ribA activator antisense (Chappell et al., 2015)	Sense plasmid backbone	ttgacagctagctcagtcctaggtataatactagtT ATTGTTGAACGCGTACCATTGATTGT AGGTCGTAACCCCAATAACGAACATT ATCTCGATACCAAAGCCGAGAAAATG GGCCATTTGCTGAACAAATAACCCTC TTGCATTGTGTAATTCATTTGCTTGC CGGAAGCAAAATAACCGGCAAGCAAA TAGTTGTTACTggatctaggaggaaggatct	3.5, B.7
JBL3403	ribA activator sense (Chappell et al., 2015)	JBL002	ttgacagctagctcagtcctaggtataatactagtT TATTTTGCTTCCGGCAAGCAAATGAA TTACACAATGCAAGAGGGTTATTTGT TCAGCAAATGGggatctcaaaGCCCGCCg aaaGGCGGGCttttttt	3.5, B.7

Table B.2 Promoters used in this study. Promoters are referred to by their iGEM BioBrick numbers, with the exception of BBa_J23119 SpeI, which is a variant of BBa_J23119 that has been mutated to include a SpeI restriction site directly before the +1 transcription start site. Sequence differences from the consensus promoter (BBa_J23119) have been highlighted in red. Relative strength data (corrected for different standards) is derived from Kelly et al. (2009) as marked by * or the Anderson catalog (http://parts.igem.org/Promoters/Catalog/Anderson) as marked by **.

Promoter name	Sequence	Strength
BBa_J23119	ttgacagctagctcagtcctaggtataatgctagc	n/a (>1)
BBa_J23119 SpeI	ttgacagctagctcagtcctaggtataatactagt	n/a (>1)
BBa_J23100	ttgacggctagctcagtcctaggtacagtgctagc	1 **
BBa_J23102	ttgacagctagctcagtcctaggtactgtgctagc	0.86 **
BBa_J23151	ttgatggctagctcagtcctaggtacaatgctagc	0.266-0.424*
BBa_J23150	tttacggctagctcagtcctaggtattatgctagc	0.098-0.161*

Table B.3 Statistical significance of comparisons in Figure 2b. P-value significance of comparisons in fold activation between different pbuE sense promoter variants. The * indicates significance by a two-sided T-test ($p<0.05$).

		Sense Promoter Variant					
		J23119 SpeI	**J23119**	**J23100**	**J23102**	**J23151**	**J23150**
Sense Promoter Variant	**J23119 SpeI**	-	-	-	-	-	-
	J23119	0.8988	-	-	-	-	-
	J23100	0.0008*	0.0010*	-	-	-	-
	J23102	0.0013*	0.0014*	0.9990	-	-	-
	J23151	0.0002*	0.0002*	0.0494*	0.0526	-	-
	J23150	7.54E-5*	5.50E-5*	0.0039*	0.0041*	0.1814	-

B.2 Supplementary Figures

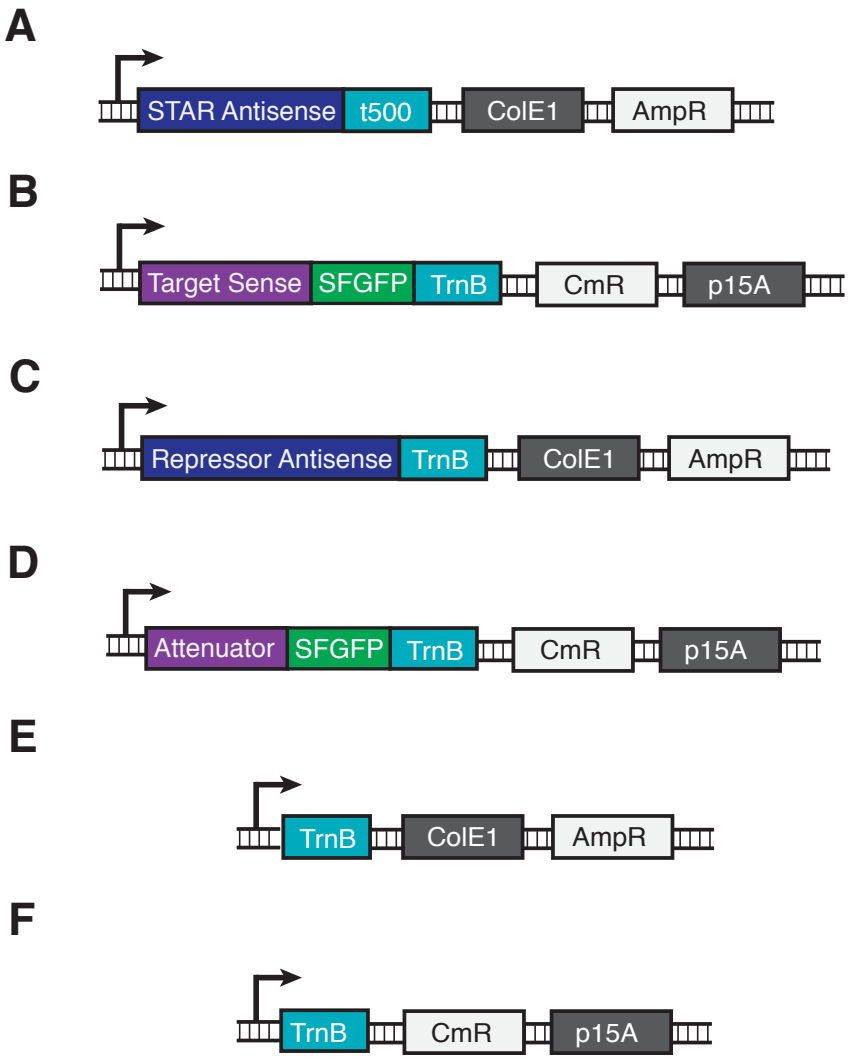


Figure B.1 Maps of DNA plasmids used in this study. (A) STAR antisense plasmid, **(B)** Target sense plasmid, **(C)** Repressor antisense plasmid, **(D)** Repressor target sense (attenuator) plasmid, **(E)** STAR/repressor antisense control plasmid, and **(F)** Target sense control plasmid. t500 and TrrnB are transcriptional terminators. ColE1 and p15A are origins of replication. AmpR and CmR are ampicillin/carbenicillin and chloramphenicol resistance cassettes, respectively.

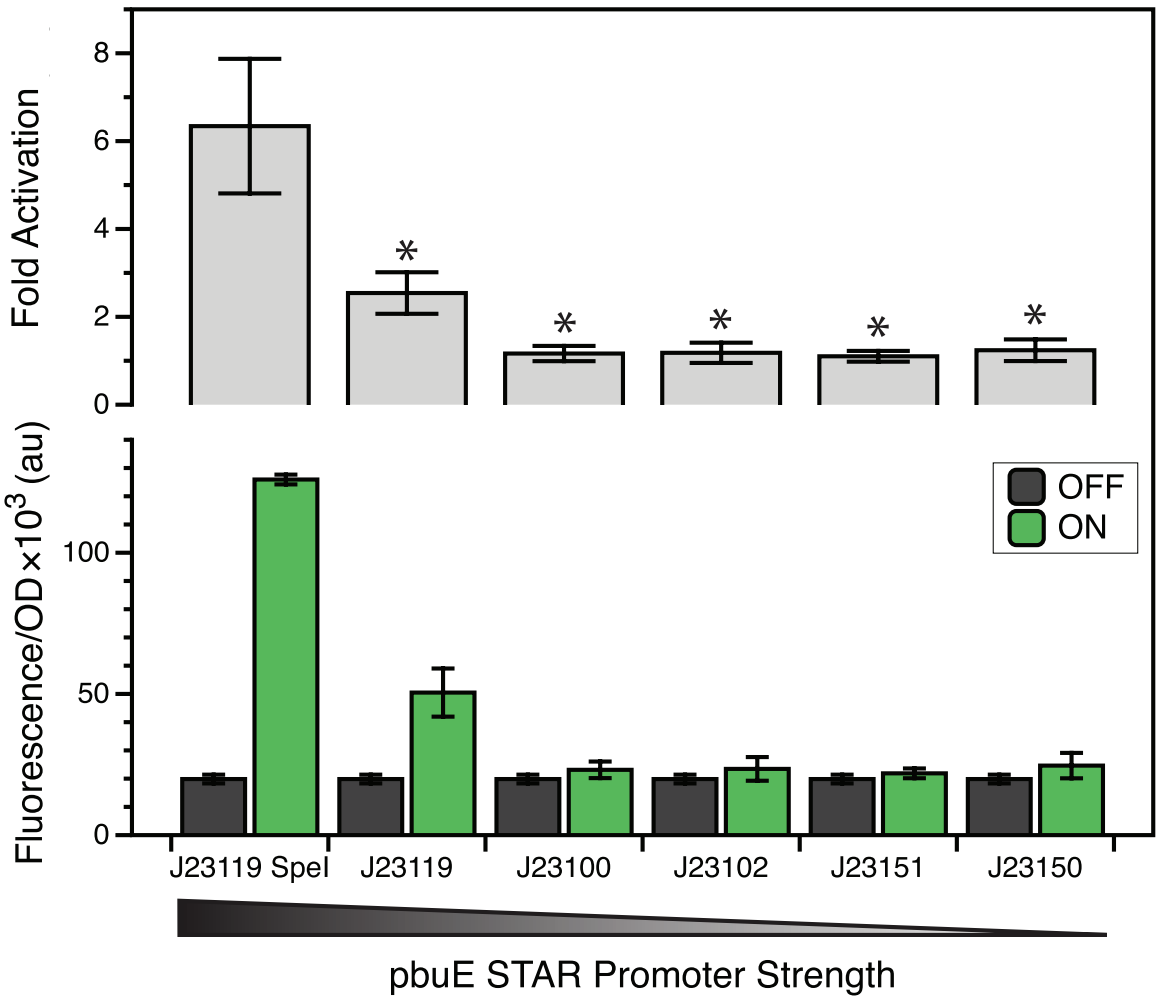


Figure B.2 Fold activation characterization for the pbuE STAR expressed from decreasing strength promoters. Normalized fluorescence divided by optical density at 600 nm (Fluorescence/OD) is plotted for the OFF case (no-STAR control plasmid) and the ON case (STAR present) on the lower plot. Promoter strength refers to the promoter used to express the STAR RNA (see Table B.2). All sense targets were expressed from the J23119 SpeI promoter. Fold activation (Fluorescence/OD ON divided by Fluorescence/OD OFF) is plotted above as a series of gray bars. Error bars represent sample standard deviation over 3 independent replicates with 3 colonies each (n=9). The * symbol indicates statistically significant ($p<0.05$) differences in fold activation between the wild-type STAR promoter configuration (J23119 SpeI) and a specific STAR promoter as determined by a two-sided T-test ($p<0.05$). It is clear from this data that weakening the STAR promoter decreases FL ON and therefore fold activation as we would expect.

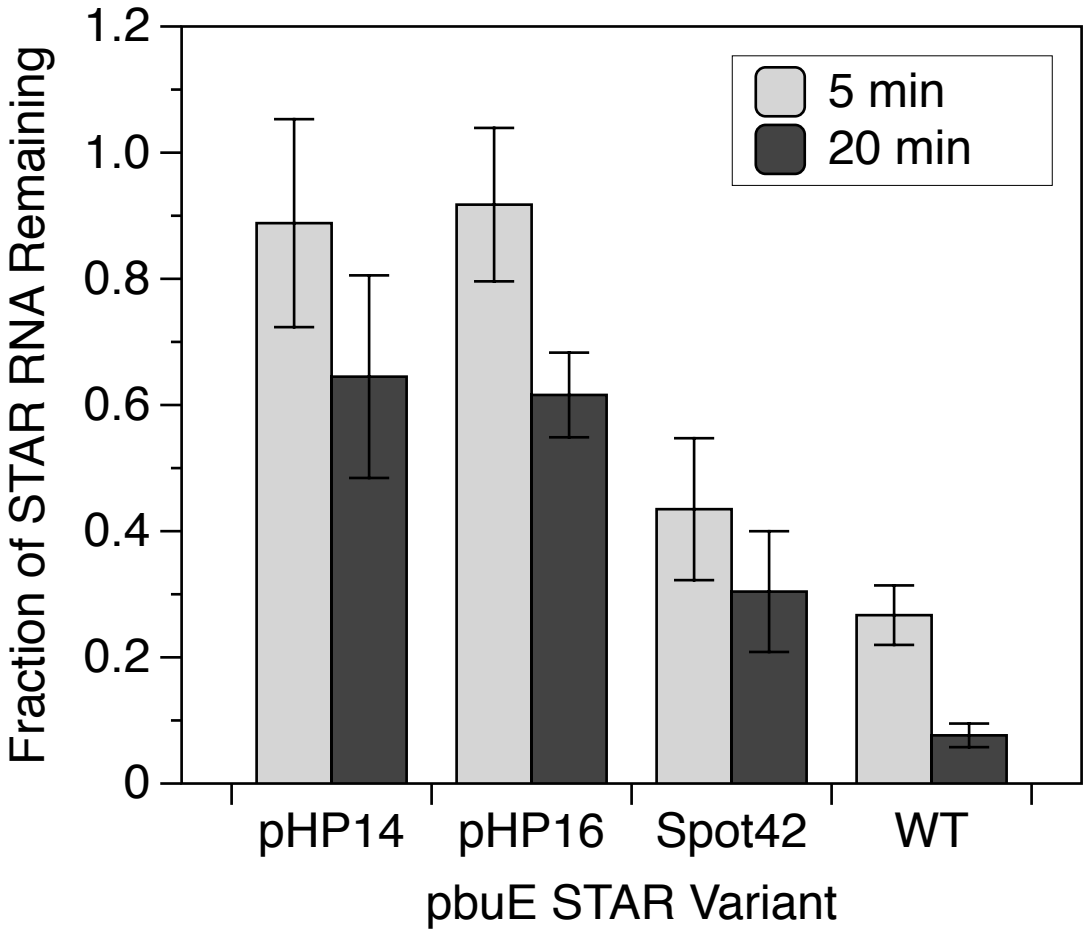


Figure B.3 pbuE STAR degradation as a function of time as determined by RT-qPCR. *E. coli* were transformed with plasmids encoding several pbuE STAR variants from Figure 3 containing RNA stability hairpins (pHP14, pHP16), an sRNA scaffold (Spot42), or neither (WT). After overnight culture, cells were subcultured for 5 hours, at which point rifampicin was added to the media to stop transcription (see B.3 Supplementary Methods). Total RNA was extracted from cells after 0, 5 and 20 minutes after rifampicin addition. Each pbuE STAR variant and 16s rRNA was quantified for each time point using RT-qPCR primers specific to these RNAs. These values were then used to calculate a normalized amount of pbuE variant RNA relative to 16s rRNA following Bernstein(Bernstein et al., 2004). Relative pbuE variant RNA amounts at the 5 and 20 minute time points were then divided by the 0 minute time point value to calculate the fraction of STAR RNA remaining according to Pfaffle (2001), see Supplementary Methods (B.3). Each RT-qPCR reaction was performed in triplicate on three biological replicates (9 reactions total for each variant). Bars represent averages over these replicates, and error bars represent the standard deviation following propagation of errors over these replicates through the normalization calculations (Pfaffl, 2001). It is clear from this analysis that the pHP14 and pHP16 hairpins stabilize the pbuE RNA to a greater extent than the Spot42 scaffold.

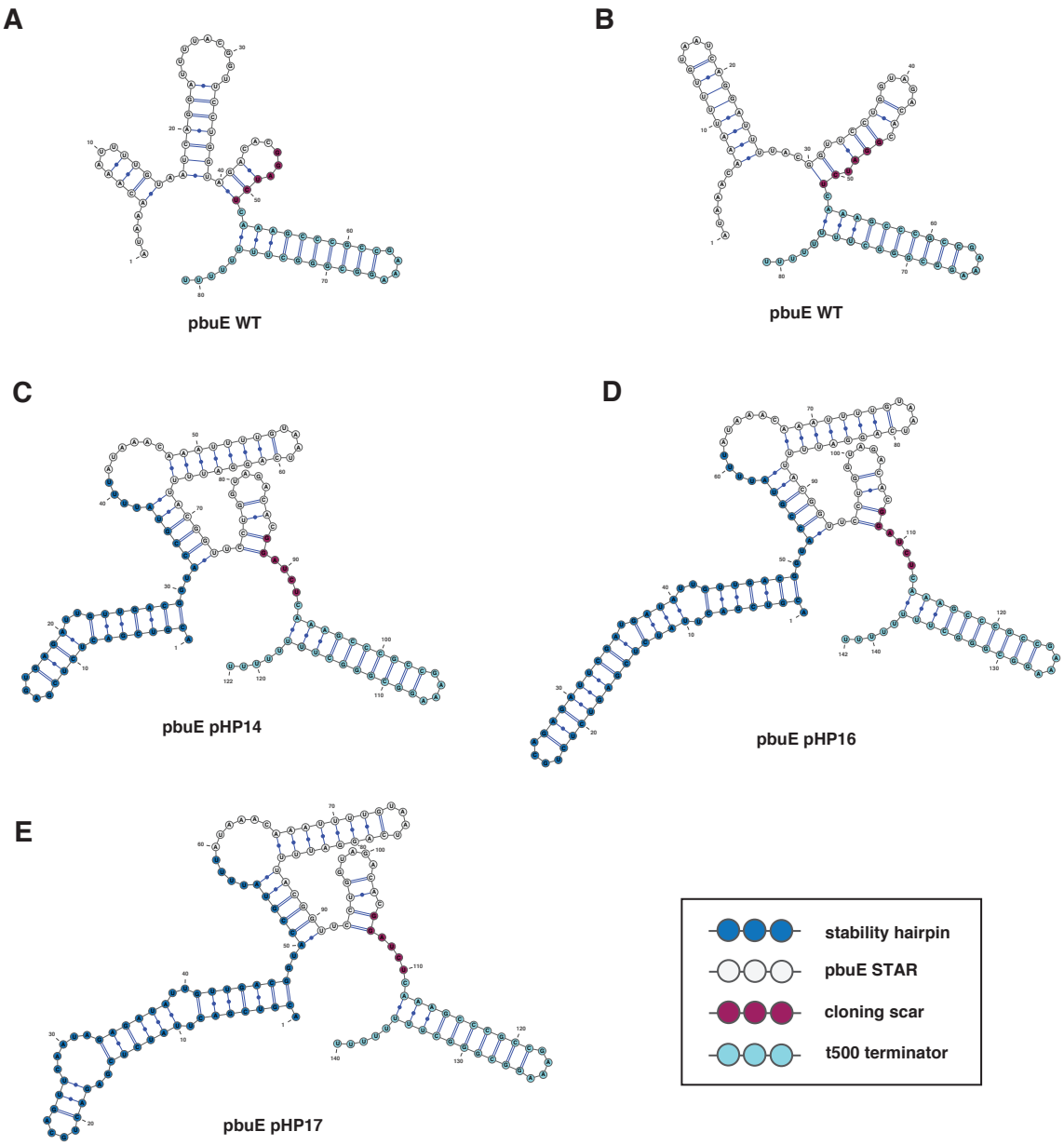


Figure B.4 Minimum free-energy (MFE) RNA fold(s) for pbuE WT STAR and stability hairpin fusions to the pbuE STAR as predicted by RNAStructure (Reuter and Mathews, 2010). **(A)** and **(B)** two MFE folds for pbuE WT. Single MFE folds for **(C)** pbuE pHP14, **(D)** pbuE pHP16, and **(E)** pbuE pHP17.

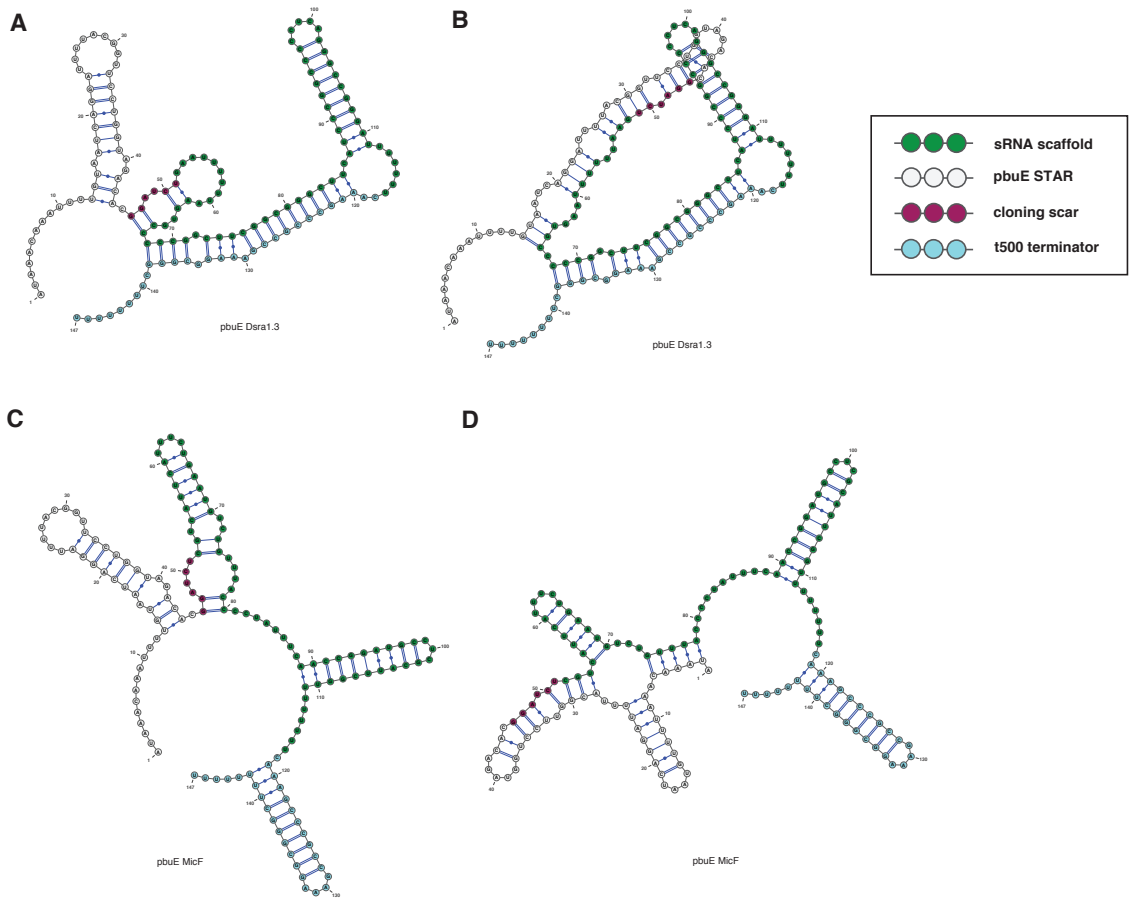


Figure B.5 Low-energy RNA folds for sRNA scaffold fusions to the pbuE STAR as predicted by RNAStructure (Reuter and Mathews, 2010). **(A)** and **(B)** pbuE Dsra1.3 fusion, **(C)** and **(D)** pbuE MicF fusion. Structures for pbuE WT are presented in Figure B.4.

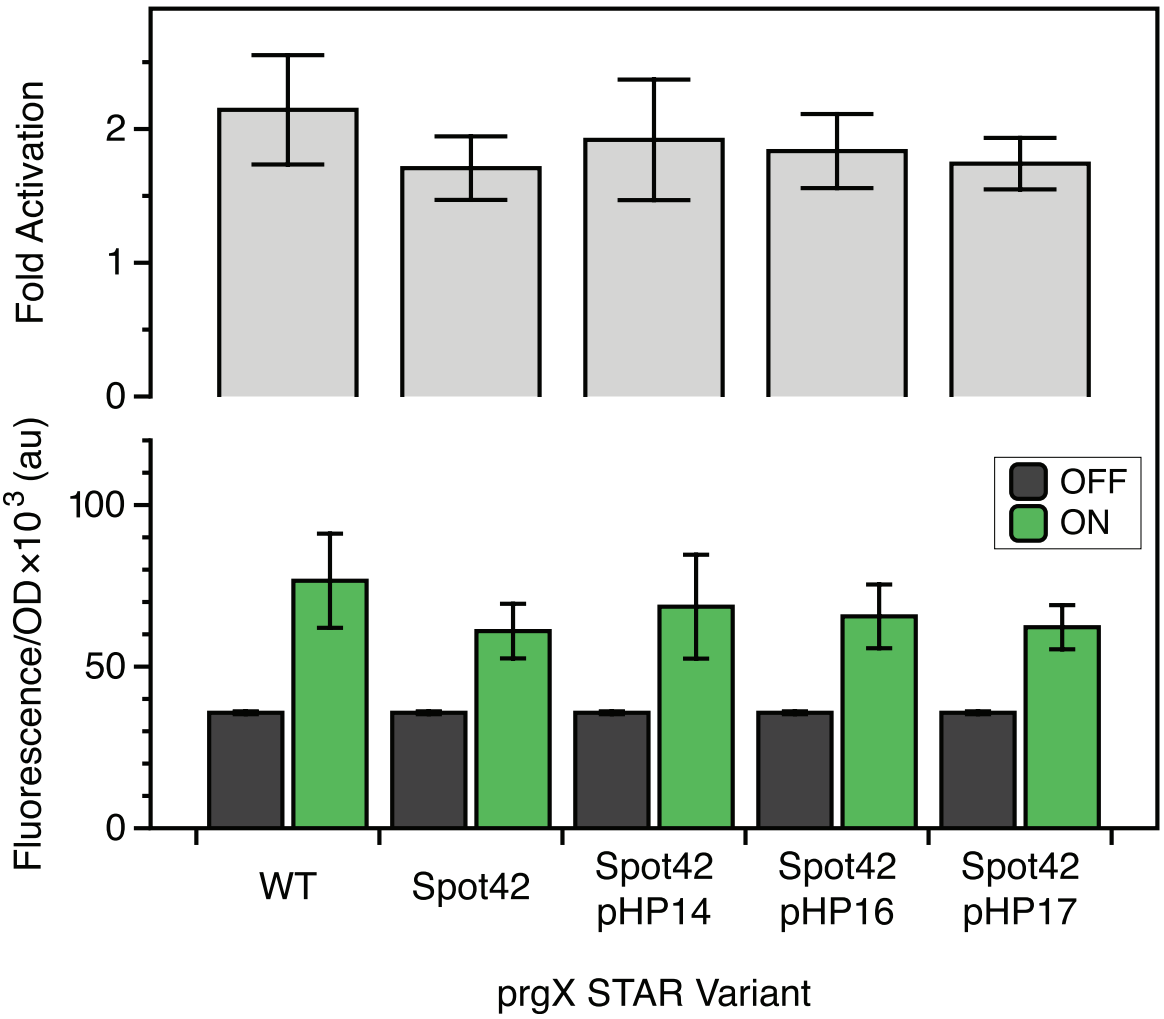


Figure B.6 Testing the addition of Spot42 to prgX antisense variants with and without stability hairpins. Normalized fluorescence divided by optical density at 600 nm (Fluorescence/OD) is plotted on the lower axes, with the dark gray bars representing the OFF level (no-STAR control plasmid) and the green bars representing the ON level (STAR present). Fold activation (Fluorescence/OD ON divided by Fluorescence/OD OFF) is plotted on the upper axis as a series of gray bars. Error bars represent sample standard deviation over 3 independent replicates with 3 colonies each (n=9).

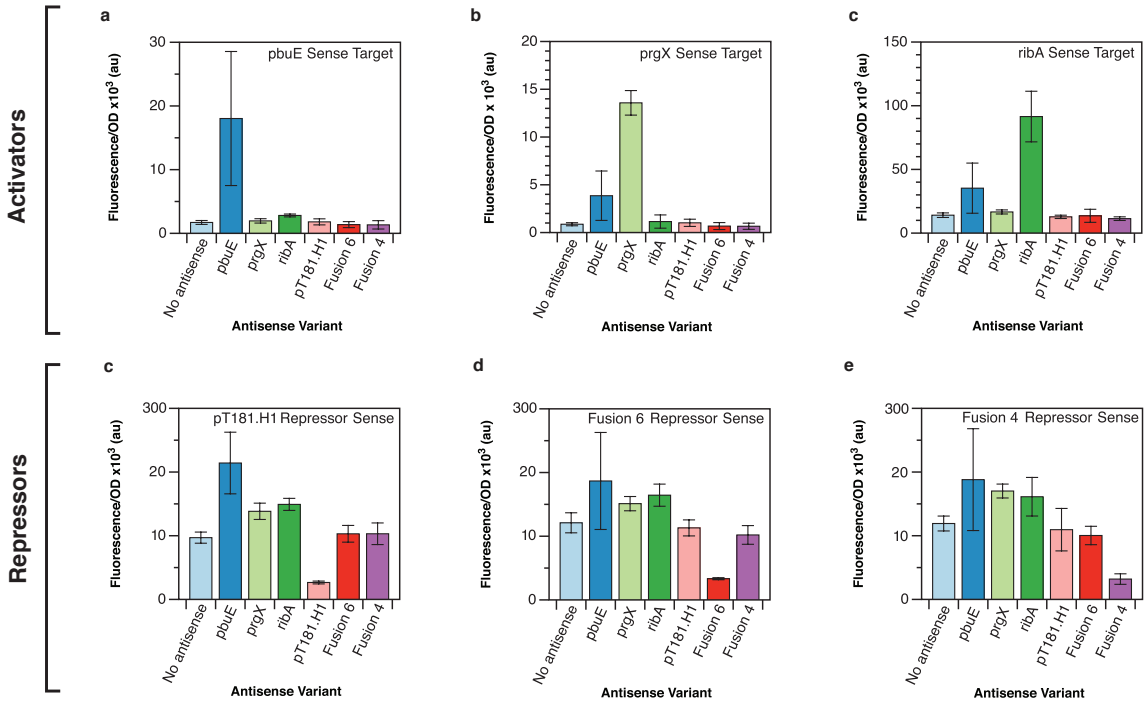


Figure B.7 *In vivo* fluorescence data from Figure 3.5 orthogonality testing. Data shows testing of optimized pbuE and prgX STAR variants against the previously reported ribA STAR (Chappell et al., 2015) and pT181.H1 (Lucks et al., 2011), Fusion 4, and Fusion 6 repressors (Takahashi and Lucks, 2013). Data is plotted as normalized fluorescence divided by optical density at 600 nm (Fluorescence/OD). Error bars represent sample standard deviation over 3 independent replicates with 3 colonies each (n=9).

B.3 Supplementary Methods

Total RNA extraction for quantitative PCR. For all extraction of total RNA for reverse transcription quantitative PCR (RT-qPCR) experiments *E. coli* strain TG1 was used. Plasmids were transformed and subsequent colonies grown overnight as described for *in vivo* bulk fluorescence measurements. For each of three biological replicates, 34 μL of a single overnight culture were added to wells containing 1666 μL (1:50 dilution) of supplemented M9 minimal media containing the selective antibiotics and grown for 5 h at the same conditions as the overnight cultures. At this point, 8.5 μL of 50 mg/mL rifampicin in DMSO were added to halt transcription. From this transcription-halted sample were taken three time-point samples of 500 μL at 0, 5, and 20 minutes after rifampicin addition. The time-point samples were pelleted by centrifugation at 13,000 rpm for 1 min. The supernatant was removed and the remaining pellet resuspended in 750 μL of Trizol reagent (Life Technologies), homogenized by repetitive pipetting, incubated at room temperature for 5 min, and then stored on ice for approximately 2 hours. Then 150 μL of chloroform (Sigma Aldrich) were added, and the samples were mixed for 15 sec and incubated at room temperature for 3 min. Following incubation, the samples were centrifuged for 15 min at 15,000 rpm at 4 °C and 200 μL of the top aqueous layer removed. One μL of glycogen (20 mg/mL) (Life Technologies) and 375 μL of chilled isopropanol were added to the aqueous phase, which was incubated at room temperature for 10 min and centrifuged for 15 min at 15,000 rpm at 4 °C. Following centrifugation the isopropanol was carefully removed from the total RNA/glycogen pellets, which were then washed in 600 μL of chilled 70 % ethanol (EtOH) and centrifuged for 2 min at 15,000 rpm at 4 °C. The EtOH was removed and the tubes centrifuged for another 2 min at 15,000 rpm at 4 °C followed by additional ethanol removal and drying at room temperature for 10 minutes to evaporate any additional ethanol. Pellets were resuspended in 20 μL of RNase free double-distilled water (ddH_20).

DNase treatment of total RNA for qPCR. Purified total RNA samples were quantified by the Qubit Fluorometer (Life Technologies) and were diluted to a concentration of 10 ng/μL

in a total of 10 μL RNase free ddH_20 and digested by Turbo DNase (Life Technologies) for 1 h at 37 °C using 3 μL Turbo DNAse per 50 μL reaction with buffer concentration as specified by manufacturer. After digestion 150 μL of RNase free ddH_20 and 200 μL phenol/chloroform (Acros Organics) were added, and the samples were vortexed for 10 sec and incubated for 3 min at room temperature before centrifugation for 10 min at 15,000 rpm at 4 °C. After centrifugation, 190 μL of the top aqueous layer was carefully removed, 190 μL of chloroform added, and the samples were vortexed for 10 sec, incubated for 3 min at room temperature and centrifuged for 10 min at 15,000 rpm at 4 °C. After centrifugation, 170 μL of the top aqueous layer was carefully removed, 170 μL of chloroform added, and the samples were vortexed for 10 sec, incubated for 3 min at room temperature and centrifuged for 10 min at 15,000 rpm at 4 °C. After centrifugation, 120 μL of the top aqueous layer was carefully removed and added to 1 μL glycogen, 360 μL of chilled 100 % EtOH and 12 μL of 3 M sodium acetate pH 5.5. Samples were vortexed for 10 sec and stored at -80 °C for 1 h. Samples were then centrifuged for 30 min at 15,000 rpm at 4 °C. Supernatant was removed and the pellets washed in 600 μL of chilled 70 % EtOH. Samples were then centrifuged for 2 min at 15,000 rpm at 4 °C and the EtOH removed. Samples were re-centrifuged for 2 min at 15,000 rpm at 4 °C and residual EtOH removed, pellets air-dried for 10 min and eluted in 10 μL RNase free ddH_2O.

Normalization of total RNA, reverse transcription and qPCR measurements. Each sample was quantified by Qubit Fluorometer and the sample diluted to 1 ng/μL of total RNA in 10 μL RNase free ddH_20. Then each DNase-treated total RNA dilution was subjected to reverse transcription of the target pbuE STAR RNA and of 16s rRNA (for normalization). In the reverse transcription reaction, 1 μL of 1 ng/μL total RNA, 0.5 μL of 2 μM reverse transcription primer (Integrated DNA Technologies: 16s rRNA RT primer = TAAGGAGGTGATCCAACCG; pbuE STAR RT primer = AGATCCGTGTCTACCAGGAA), 0.5 μL of 10 mM of dNTPs (New England Biolabs) and RNase free ddH_20 up to 6.5 μL were incubated for 5 min at 65 °C and cooled on ice for 5 min. 0.25 μL of Superscript III reverse transcriptase (Life technologies), 0.5 μL of 100 mM Dithiothreitol (DTT), 2 μL 5x first-strand buffer (Life technologies), and 0.5 μL RNaseOUT (Life Technologies) and 0.25 μL

RNase free H_2O up to 3.5 μL were then added, incubated at 55 °C for 1 h, 75 °C for 15 min and then stored at -20 °C. qPCR was performed on the reverse transcription products using 5 μL of Maxima SYBR green qPCR master mix (Thermo Scientific), 1 μL of cDNA and 0.5 μL of 2 μM pbuE qPCR primers (Integrated DNA Technologies, FWD = ATAAACAAATTTTGTAATCAGGATTTTACGG, REV = AGATCCGTGTCTACCAGGAA) or 16s qPCR primers (Integrated DNA Technologies, FWD =GTCAGCTCGTGTTGTGAAATG, REV = CCCACCTTCCTCCAGTTTATC) and RNase free ddH_2O up to 10 μL. Control reactions using 0.1 ng/μL DNase-treated total RNA to check for background genomic DNA contamination were also assembled, and primer efficiency was quantified by a five-point standard curve using purified PCR product corresponding to the sequence generated by reverse-transcription. It was shown that the qPCR primer sets had primer efficiencies between 71-108%. A ViiA 7 real-time PCR machine (Applied Biosystems) was used for data collection using the following PCR program: 50 °C for 2 min, 95 °C for 10 min followed by 30 cycles of, 95 °C for 15 sec and 53 °C for 1 min. A MicroAmp EnduraPlate Optical 384-well plate (Applied Biosystems) and an optically clear seal (Applied Biosystems) were used for all measurements. Results were analyzed using the ViiA 7 software (Applied Biosystems) to identify C_T values for each sample, which were then further analyzed according to the description below. All cDNA samples were measured in triplicate, and non-template controls were run in parallel to check for contamination and non-specific amplification or primer dimers. These non-template controls showed insignificant amounts of contamination, with the highest level being 399 times lower than the corresponding measured sample. In addition, qPCR was performed on total RNA samples to confirm no DNA plasmid was detected under conditions used. Melting curve analysis was performed to confirm that only a single product was amplified.

Data Analysis

Calculations of the relative abundance of pbuE variant STAR RNA remaining at each time point were made following the method laid out by Pfaffl (2001) which accounts for differences in amplification efficiency between target and reference. Specifically, for each time point t_i, we used this analysis to calculate a normalized pbuE variant RNA abundance,

N(i), normalized to the abundance of 16s rRNA at that time point following Bernstein (2004):

$$N(i) = \frac{E_{ref}^{c_T^{ref}(i)}}{E_{target}^{c_T^{target}(i)}}$$

where E_{target} is the primer efficiency for the pbuE RT-qPCR reaction, E_{ref} is the primer efficiency for the 16s rRNA RT-qPCR reaction, $C_T^{target}(i)$ is the C_T count for the pbuE variant sample and $C_T^{ref}(i)$ is the C_T count for the 16s rRNA sample. Following Pfaffl (2001) we then calculated the fraction of pbuE STAR variant remaining as N(i)/N(0) to obtain the final formula:

$$\frac{N(i)}{N(0)} = \frac{E_{target}^{c_T^{target}(0) - c_T^{target}(i)}}{E_{ref}^{c_T^{ref}(0) - c_T^{ref}(i)}}$$

Each reaction was run in triplicate, giving a standard deviation over C_T values which were used to calculate an error in N(i)/N(0) using standard propagation of error techniques. These errors were then combined over the three full biological replicates of the time course experiment.

B.4 Bibliography

Bernstein, J.A., Lin, P.-H., Cohen, S.N., and Lin-Chao, S. (2004). Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. Proc. Natl. Acad. Sci. *101*, 2758–2763.

Chappell, J., Takahashi, M.K., and Lucks, J.B. (2015). Creating small transcription activating RNAs. Nat. Chem. Biol. *11*, 1–9.

Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. J. Biol. Eng. *3*, 4.

Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D., and Arkin, A.P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc. Natl. Acad. Sci. *108*, 8617–8622.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Takahashi, M.K., and Lucks, J.B. (2013). A modular strategy for engineering orthogonal chimeric RNA transcription regulators. Nucleic Acids Res. *41*, 7577–7588.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR ENGINEERING

LIGAND-RESPONSIVE RNA ACTIVATORS

C.1 Supplementary Figures

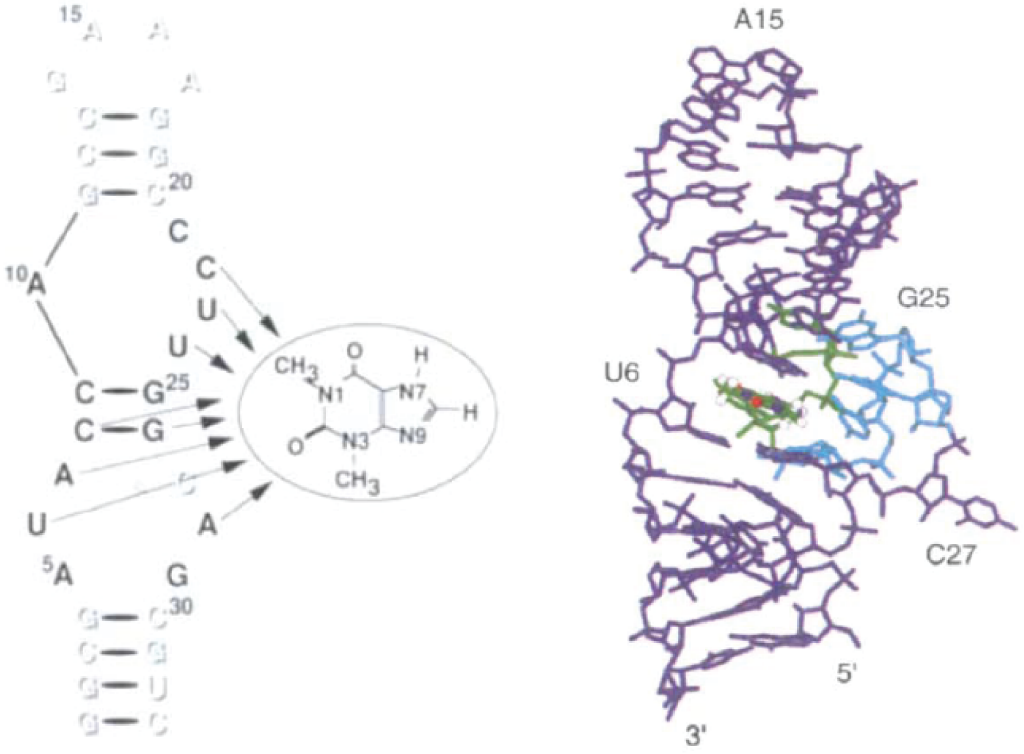


Figure C.1 Structure of the theophylline aptamer complexed with theophylline as determined by NMR spectroscopy. Theophylline binds nucleotides within an internal loop region. Reprinted by permission from Macmillan Publishers Ltd: [Nature structural biology] Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P., and Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nat. Struct. Biol. *4*, 644–649., copyright 1997.

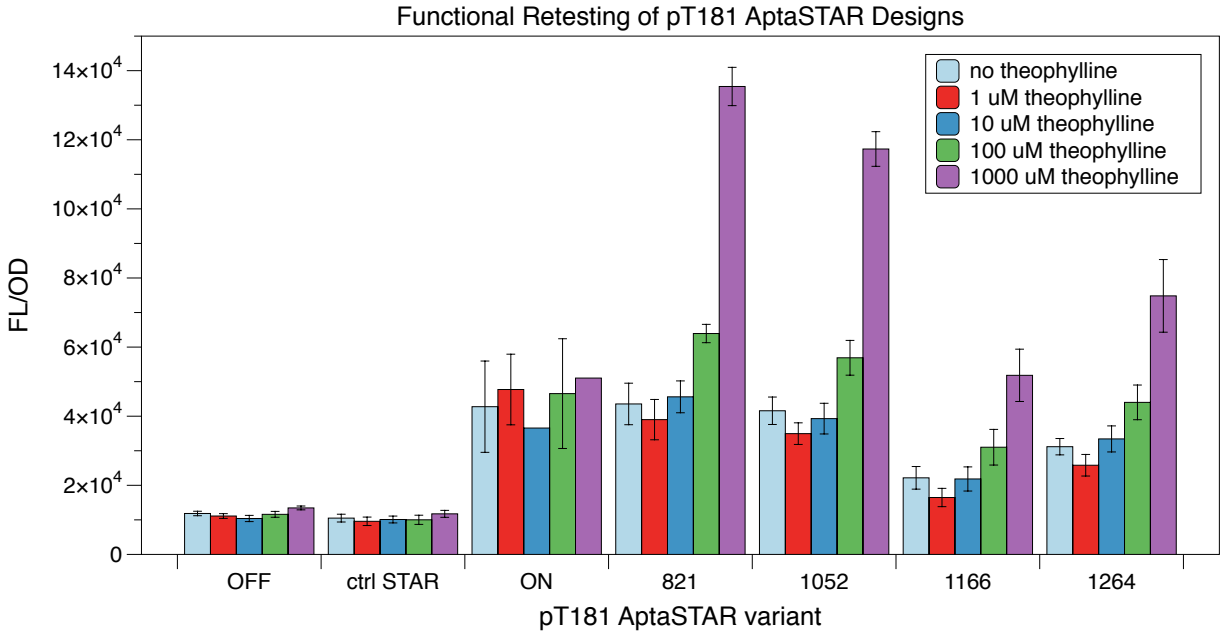


Figure C.2 Confirmation of theophylline-sensitive activation of pT181 aptaSTARs. Repeated i*n vivo* testing of computationally designed pT181 aptaSTARs confirms theophylline responsive activation in the presence of high concentrations of theophylline.

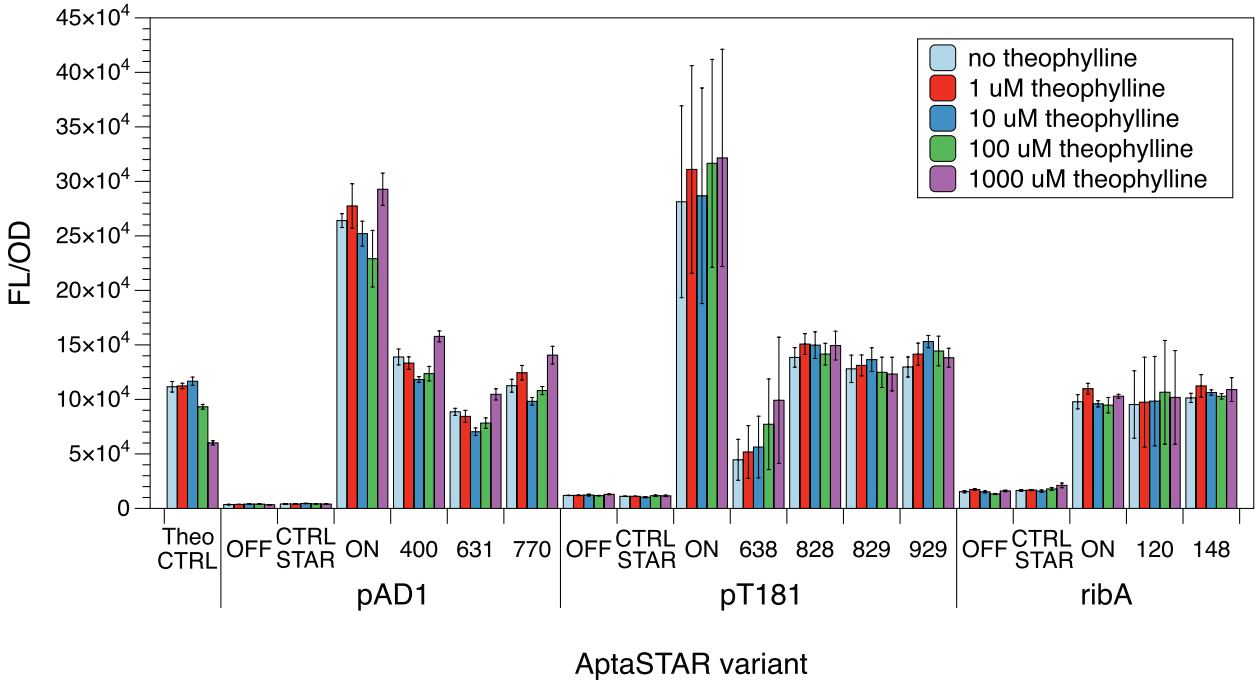


Figure C.3 Testing theophylline-sensitivity *in vivo* of computationally designed aptaSTARs with long linker regions. Theophylline aptaSTAR variants based on the pAD1, pT181, and ribA systems and designed allowing for linker sequences up to 60 nucleotides were tested alongside a theophylline-sensitive repressor (Theo CTRL), a non-cognate STAR (CTRL STAR), and the original non-theophylline-sensitive STARs (ON).

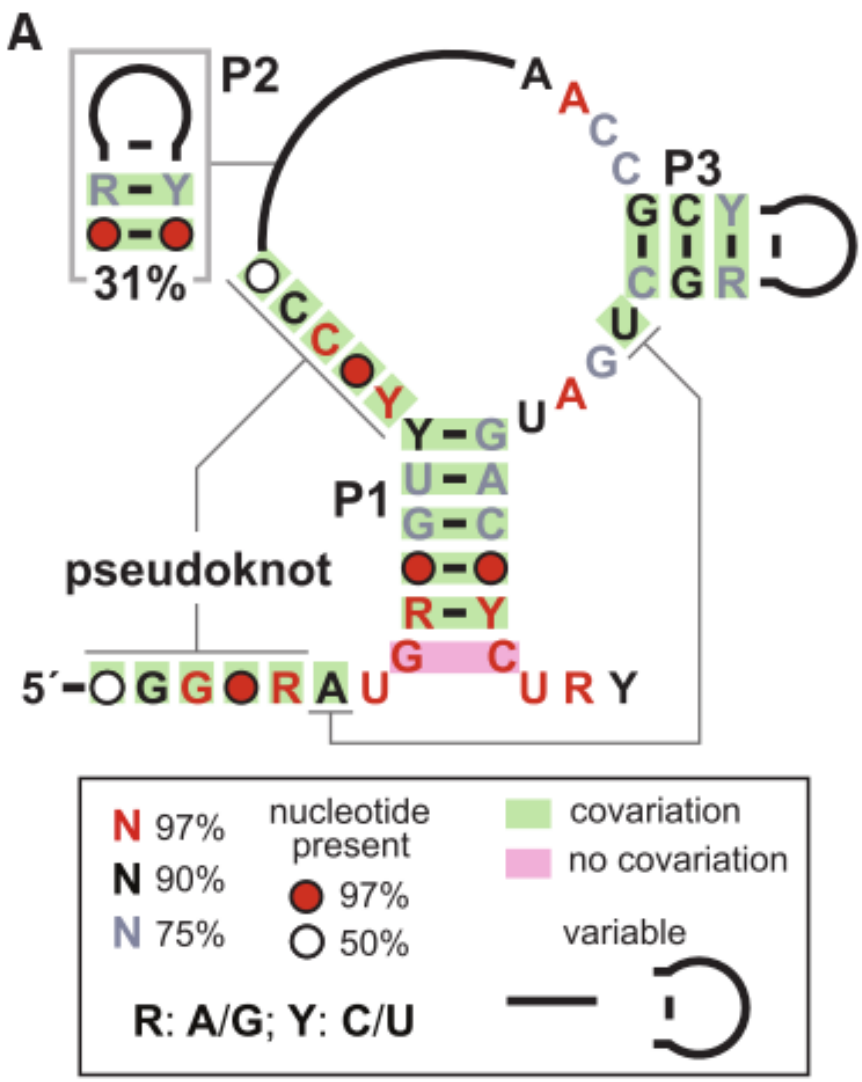


Figure C.4 Structure of the fluoride aptamer. The basic conserved structure includes two hairpin stems (P1 and P3), as well as an important pseudoknot interaction. From Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. Science. *335*, 233–235. Reprinted with permission from AAAS.

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHARACTERIZING A NATURALLY-OCCURRING RNA THERMOMETER (CHAPTER 5)

D.1 Supplementary Tables

Table D.1 Plasmids used in this study.
Color code: yellow-promoter, ---insertion sequence, green-SFGFP light blue-terminator, dark blue-chloramphenicol resistance marker, and grey-origin of replication (p15a)

Plasmid number	Description	Sequence
n/a	p15a backbone	gaattctaaagatctttgacagctagctcagtcctaggtatatactagt---------- ATGagcaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtga tgttaatgggcacaaattttctgtccgtggagagggtgaaggtgatgctacaaacggaaaa ctcacccttaaatttatttgcactactggaaaactacctgttccgtggccaacacttgtcacta ctctgacctatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttc aagagtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacggga cctacaagacgcgtgctgaagtcaagtttgaaggtgatacccttgttaatcgtatcgagttaa agggtattgattttaaagaagatggaaacattcttggacacaaactcgagtacaactttaac tcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaa attcgccacaacgttgaagatggttccgttcaactagcagaccattatcaacaaaatactcc aattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgtcctttcg aaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggatt acacatggcatggatgagctctacaaataaggatctgaagcttgggcccgaacaaaaactc atctcagaagaggatctgaatagcgccgtcgaccatcatcatcatcattgagtttaaac ggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatca gaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtccca cctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctc cccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaaga ctgggcctttcgttttatctgttgtttgtcggtgaactggatccttactcgagtctagactgcag ttgatcgggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgt attttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactg gatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcag ttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaa agaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctc atccggaatttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcaccct tgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgac gatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcc tatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcac cagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaa tattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgt

		gatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcaggg cggggcgtaatttgatatcgagctcgcttggactcctgttgatagatccagtaatgacctcag aactccatctggatttgttcagaacgctcggttgccgccgggcgtttttattggtgagaatcc aagcctccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacag ggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaa gtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccag tggcttctgtttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaa aagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatga gggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcag aatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgac tgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaaga tacttaacagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccccct gacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataa agataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggta ggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgcct tatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagc agccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaact gaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggt agctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagagcaagaga ttacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctag atttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgt aattctcatgtttgacagcttatcatcgataagcttccgatggcgcgccgagaggctttacac tttatgcttccggct
JBL 3048	WT *agsA* in p15a backbone	Insertion sequence: ggacaagcaatgcttgccttgatgttgaacttttgaatagtgattcaggaggttaatg
JBL 3057	*agsA* G21C mutant in p15a backbone	Insertion sequence: ggacaagcaatgcttgccttCatgttgaacttttgaatagtgattcaggaggttaatg
JBL 3070	*agsA* A29C mutant in p15a backbone	Insertion sequence: ggacaagcaatgcttgccttgatgttgaCcttttgaatagtgattcaggaggttaatg
JBL 001	empty backbone control	gaattctaaagatcctaactcgagtaatgagaagcttgggcccgaacaaaaactcatctca gaagaggatctgaatagcgccgtcgaccatcatcatcatcattgagtttaaacggtctc cagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacg cagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgac cccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatg cgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggc ctttcgttttatctgttgtttgtcggtgaactggatccttactcgagtctagactgcagttgatcg ggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttg agttatcgagatttcaggagctaaggaagctaaaatggagaaaaaaatcactggatatac caccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctca

		atgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaa ataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccgga atttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacac cgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccg gcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccct aaagggtttattgagaatatgttttcgtctcagccaatccctgggtgagtttcaccagttttg atttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatac gcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggctt ccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcg taatttgatatcgagctcgcttggactcctgttgatagatccagtaatgacctcagaactccat ctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagcctcc gatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacacca ggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgg gtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgt ttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccg ccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtca gtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtga tacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcga gcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaac agggaagtgagagggccgcggcaaagccgtttttccataggctccgccccccctgacaagca tcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccag gcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcat tccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcg ctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggta actatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactgg taattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggaca agttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagaga accttcgaaaaaccgccctgcaaggcggtttttttcgttttcagagcaagagattacgcgcag accaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgca atttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgt ttgacagcttatcatcgataagcttccgatggcgcgccgagaggctttacactttatgcttcc ggct
JBL 2348	gene expression control plasmid	gaattctaaagatctttgacagctagctcagtcctaggtatatactagtgggagaaagaag aagagatgcggaagagagaaatataacacaacaatacgtatatttatatcttccgcaaaga agaggagaggaagaatgaaatatacgagacaacaaacaataaacagcATGagcaaag gagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggca caaattttctgtccgtggagagggtgaaggtgatgctacaaacggaaaactcacccttaaat ttatttgcactactggaaaactacctgttccgtggccaacacttgtcactactctgacctatgg tgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatg cccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggacctacaagacgc gtgctgaagtcaagtttgaaggtgataccttgttaatcgtatcgagttaaagggtattgattt taaagaagatggaaacattcttggacacaaactcgagtacaactttaactcacacaatgtat acatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacg ttgaagatggttccgttcaactagcagaccattatcaacaaaatactccaattggcgatggc cctgtccttttaccagacaaccattacctgtcgacacaatctgtcctttcgaaagatcccaac

		gaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatg gatgagctctacaaataaggatcttagcataacccccttggggcctctaaacgggtcttgagg ggtttttttgtggatccttactcgagtctagactgcagttgatcgggcacgtaagaggttccaa ctttcaccataatgaaataagatcactaccgggcgtatttttttgagttatcgagatttcagga gctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaat ggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccg ttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccgg cctttattcacattcttgcccgcctgatgaatgctcatccggaatttcgtatggcaatgaaaga cggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactga aacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcg caagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatg ttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatgg acaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctg atgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgctta atgaattacaacagtactgcgatgagtggcagggcggggcgtaatttgatatcgagctcgct tggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgct cggttgccgccgggcgtttttttattggtgagaatccaagcctccgatcaacgtctcattttcgc caaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaa gtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttact gatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcct gttcagctactgacggggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagc ggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggc aggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgctt cctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaa cggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccg cggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctca aatcagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggct ccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtt tgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaac ccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagtta gtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcc tccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccct gcaaggcggtttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaa gatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagc acctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataa gcttccgatggcgcgccgagaggctttacactttatgcttccggct

D.2 Supplementary Figures

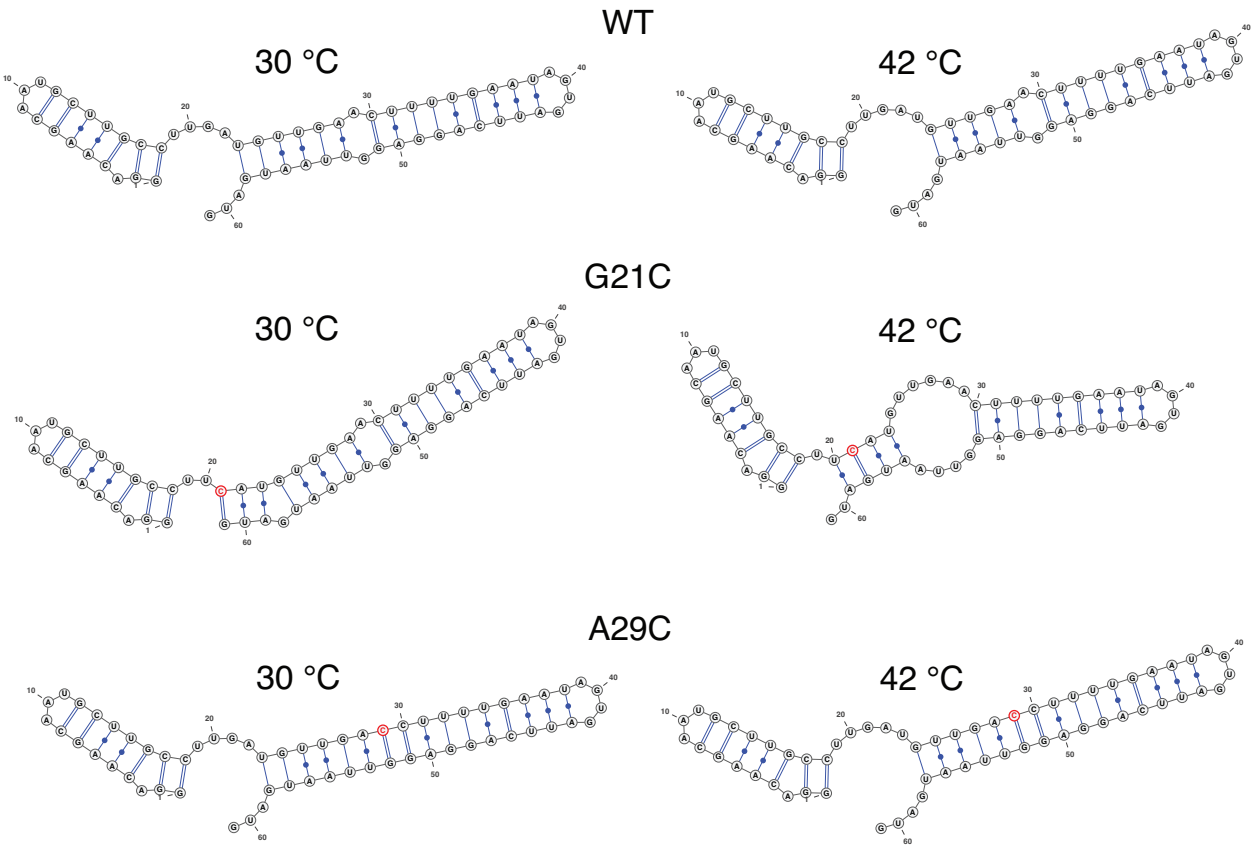


Figure D.1 Unconstrained predictions of *agsA* construct structure. Predicted *agsA* variant structures at 30 °C and 42 °C in the absence of SHAPE reactivity constraints. Mutations are highlighted in red.

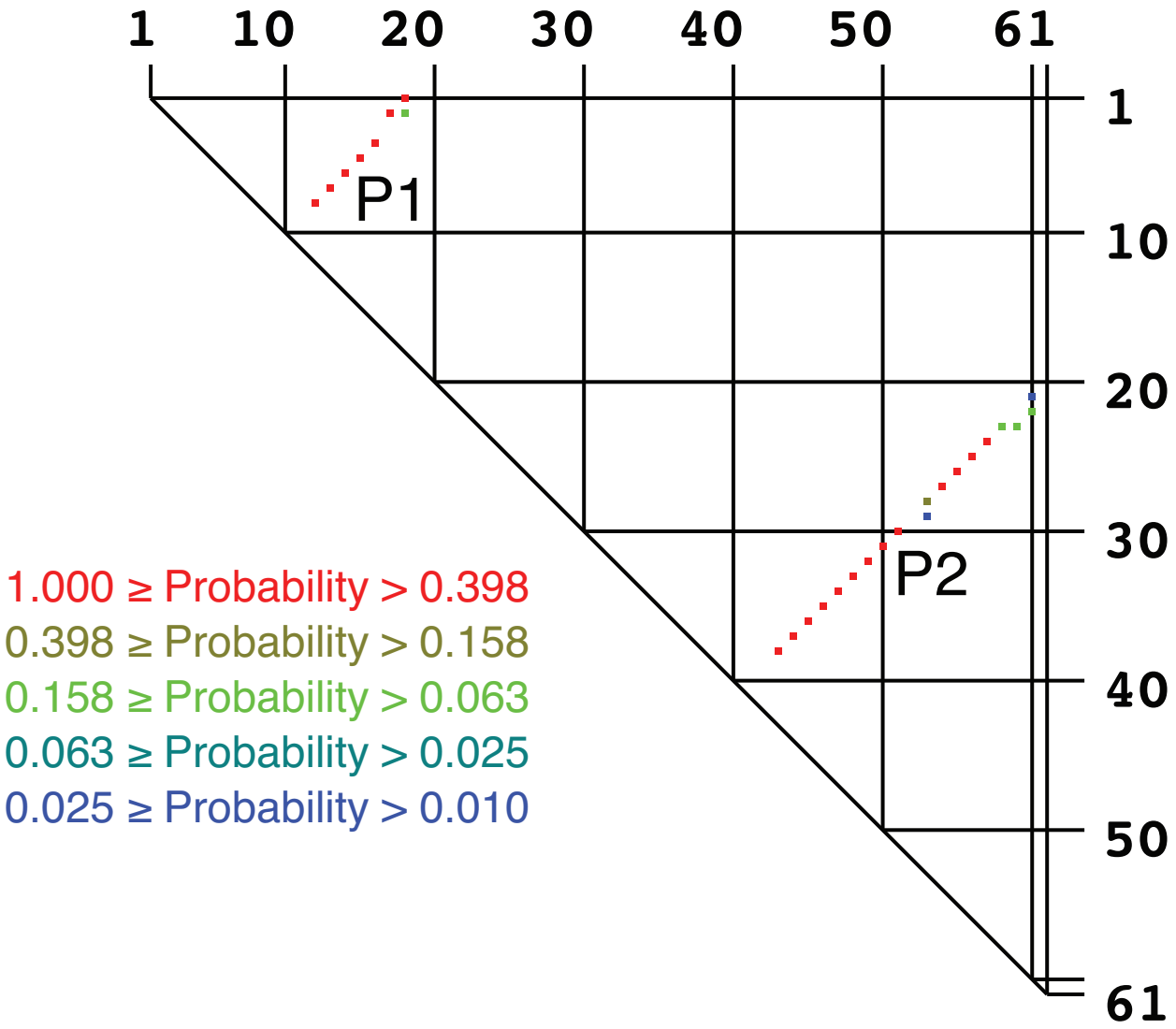


Figure D.2 *agsA* WT pairing probability matrix at 30 °C *in vivo.* Pairing probability matrix for *agsA* WT 5' UTR as predicted using RNAStructure constrained by *in vivo* SHAPE-Seq probing data. The dots represent potential base pairs that form in more than 1% of the structural ensemble, and the colors correspond to the pairing probability. The groupings of base pairs corresponding to the P1 and P2 helixes are labeled.

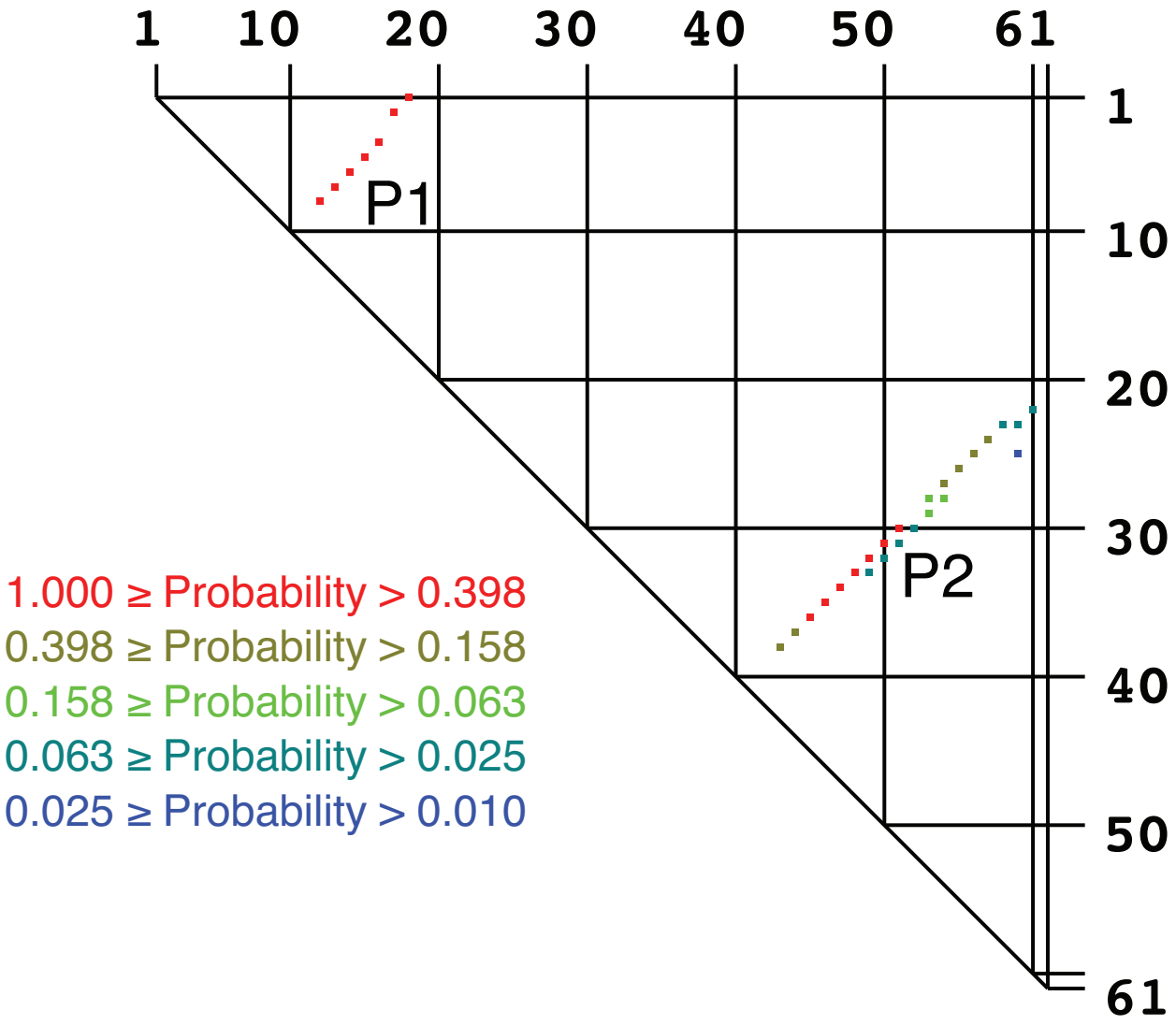


Figure D.3 *agsA* WT pairing probability matrix at 42 °C *in vivo.* Pairing probability matrix for *agsA* WT 5' UTR as predicted using RNAStructure constrained by *in vivo* SHAPE-Seq probing data. The dots represent potential base pairs that form in more than 1% of the structural ensemble, and the colors correspond to the pairing probability. The groupings of base pairs corresponding to the P1 and P2 helixes are labeled.

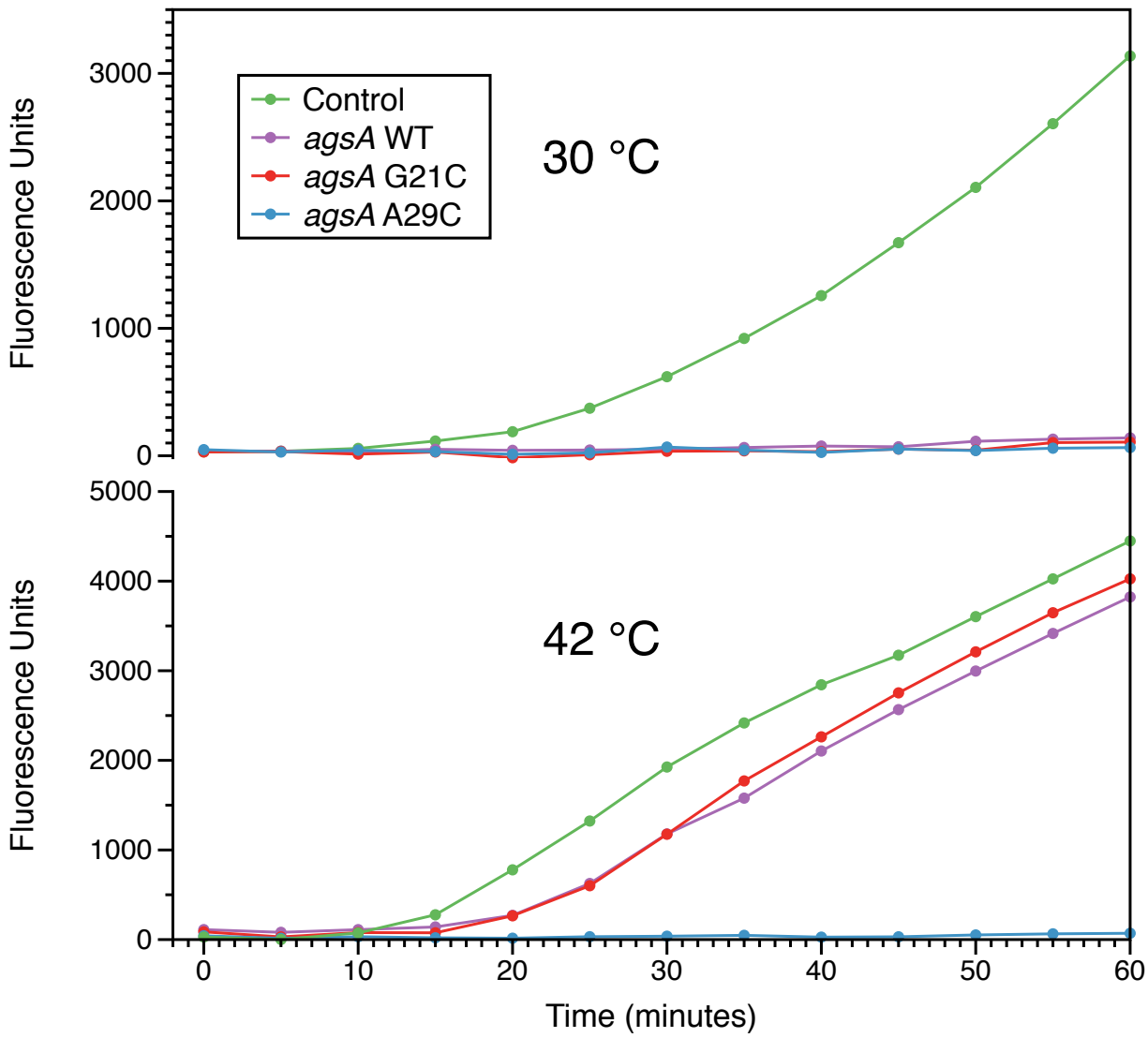
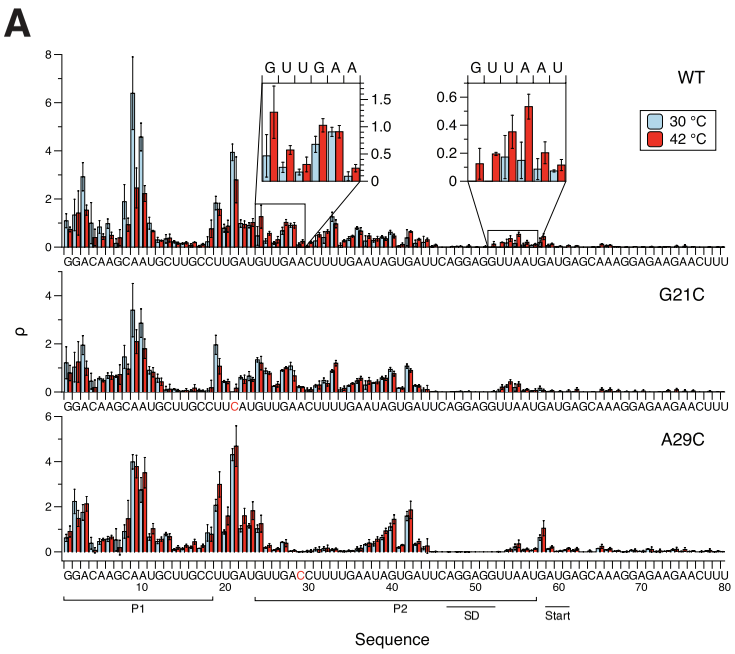
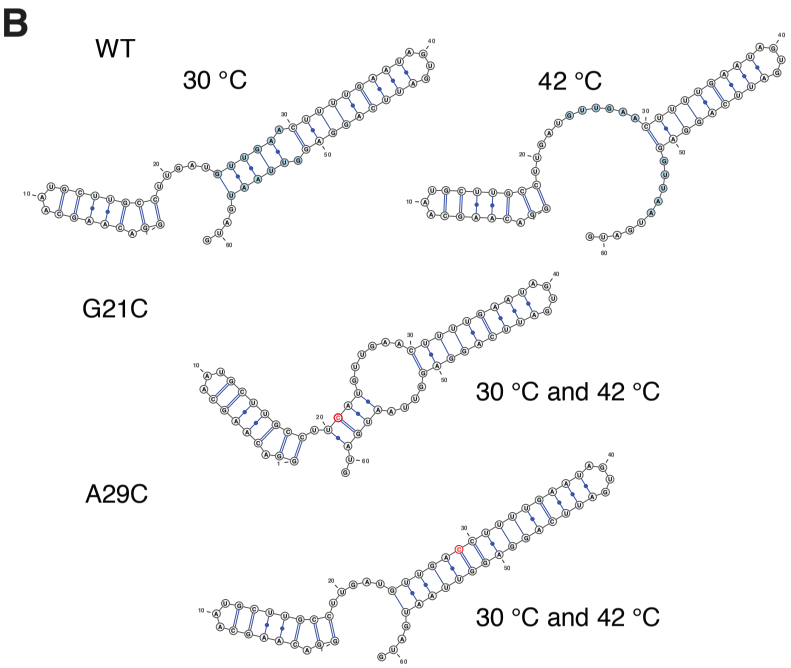


Figure D.4 PURExpress testing of *agsA* constructs from plasmid. Fluorescence traces demonstrate successful thermometer activation in the context of the PURExpress cell-free protein synthesis system.

Figure D.5 SHAPE-Seq characterization of the *agsA* thermometer *in vitro* in buffer. **(A)** SHAPE-Seq normalized reactivity plots for *agsA* WT, G21C, and A29C variants from triplicate experiments *in vitro* performed on mRNAs equilibrated at 30 °C and 42 °C. **(B)** Experimentally-derived RNA structures as predicted using the RNAStructure program with the reactivity values from (A) taken as pseudoenergy folding constraints. Blue-highlighted nucleotides undergo key structural changes and correspond to the inset boxes in (A), while red nucleotides denote variant mutations.


A
WT
G21C
A29C
30 °C
42 °C
ρ
G U U G A A
G U U A A U
GGACAAGCAAUGCUUGCCUUGAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
GGACAAGCAAUGCUUGCCUUCAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
GGACAAGCAAUGCUUGCCUUGAUGUUGACCUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
10
20
30
40
50
60
70
80
P1
P2
SD
Start
Sequence


B
WT
30 °C
42 °C
G21C
30 °C and 42 °C
A29C
30 °C and 42 °C

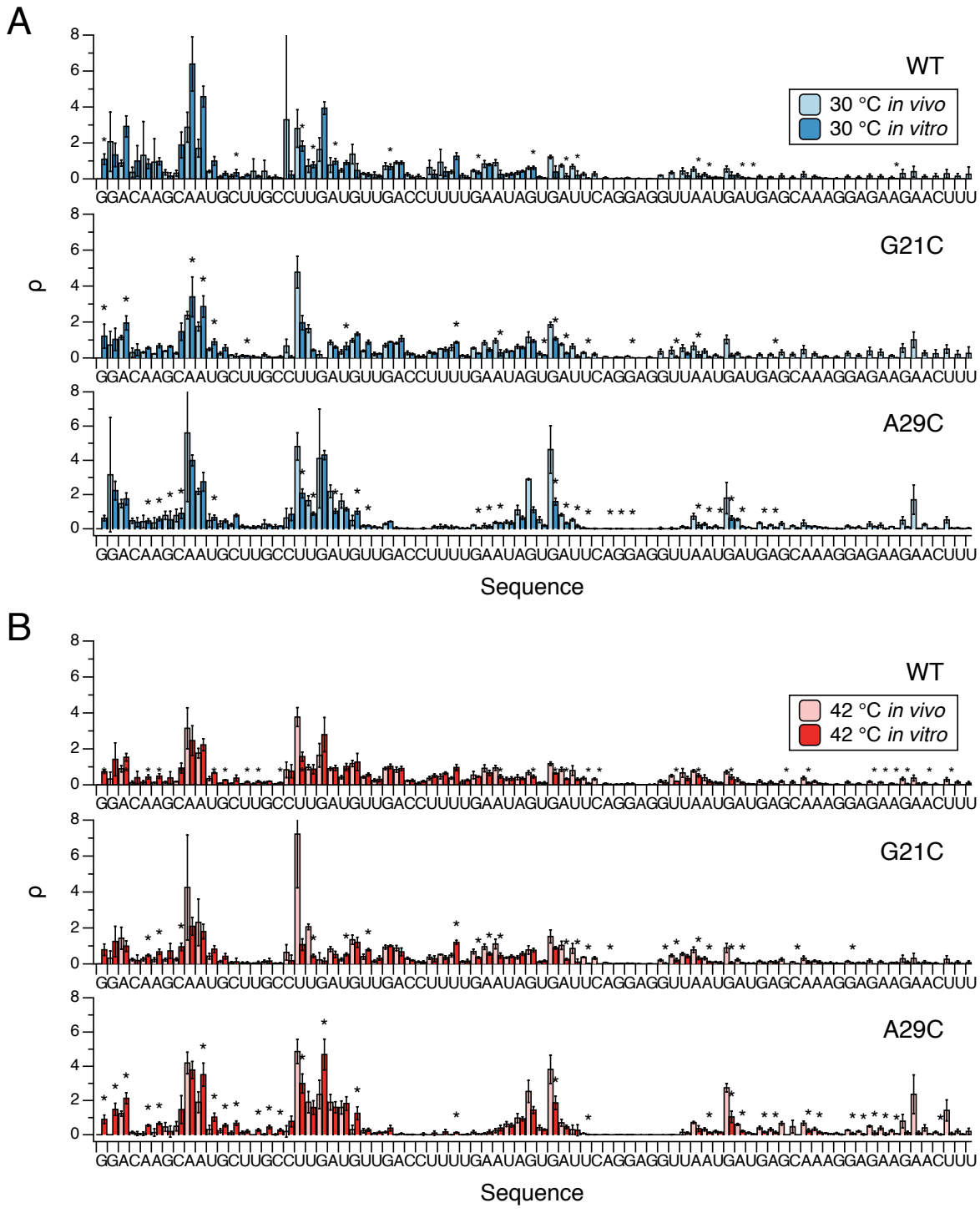
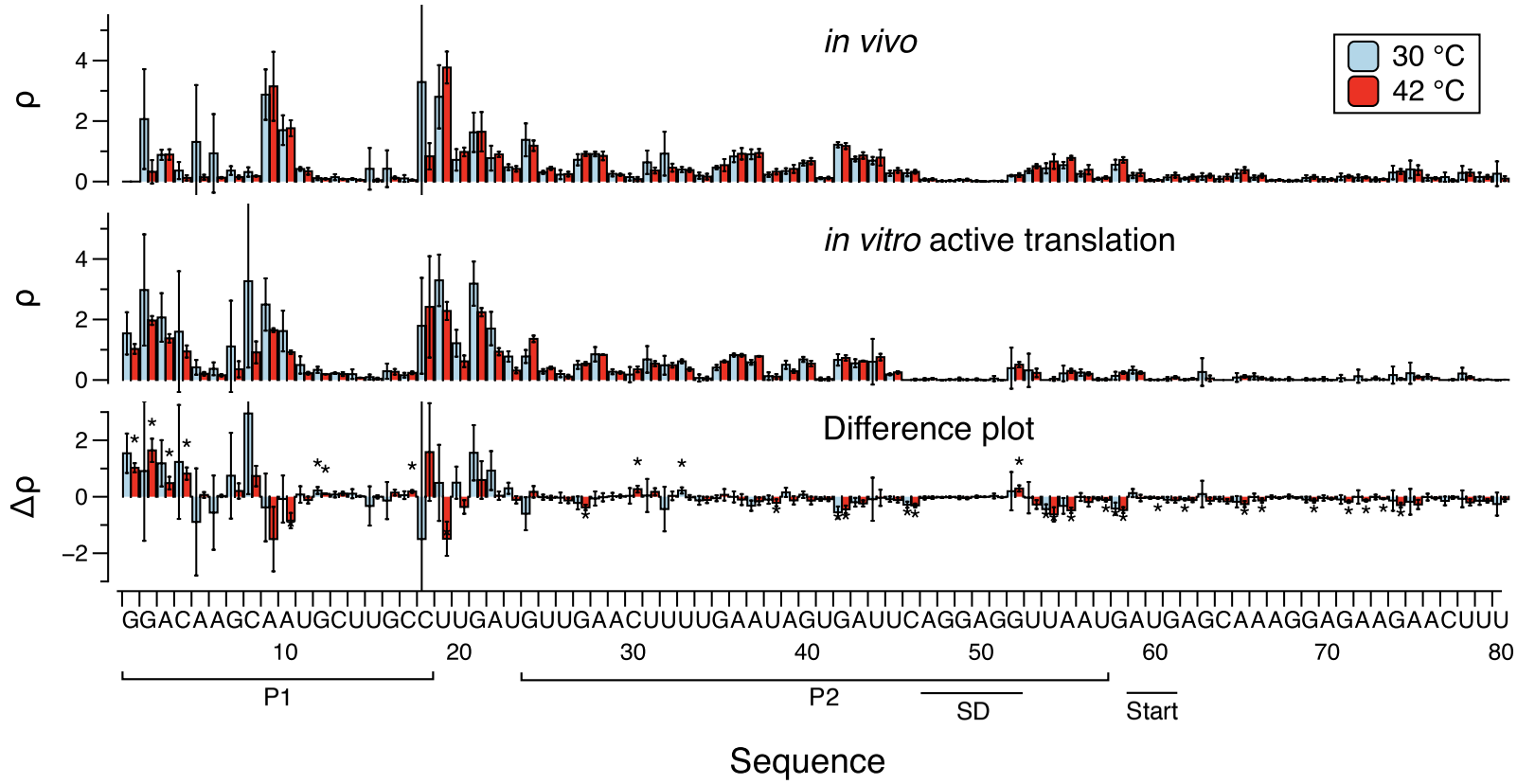


Figure D.6 The effect of active translation on mRNA reactivity. *AgsA* construct reactivity comparisons between *in vivo* and *in vitro* in buffer at 30 °C **(A)** and 42 °C **(B)**. Stars denote statistically significant reactivity differences.

Figure D.7 Comparison between SHAPE-Seq reactivities for *agsA* WT *in vivo* and *in vitro* in PURExpress with active translation. Difference plot shows *in vitro* with active translation reactivities minus *in vivo* reactivities, with stars corresponding to statistically significant differences.


in vivo
30 °C
42 °C
ρ
in vitro active translation
Difference plot
Δρ
GGACAAGCAAUGCUUGCCUUGAUGUUGAACUUUUGAAUAGUGAUUCAGGAGGUUAAUGAUGAGCAAAGGAGAAGAACUUU
10
20
30
40
50
60
70
80
P1
P2
SD
Start
Sequence

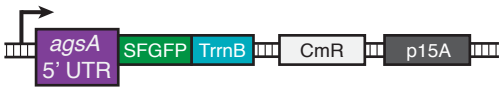


Figure D.8 Plasmid architecture of the *agsA* testing constructs. The WT or mutant *agsA* 5' UTR (purple) is fused directly after the J23119 promoter (black arrow), followed by the fluorescent reporter SFGFP (green) and the TrrnB intrinsic terminator (blue). The plasmid also harbors the chloramphenicol resistance selection marker (light gray) and the p15a origin of replication (dark gray).

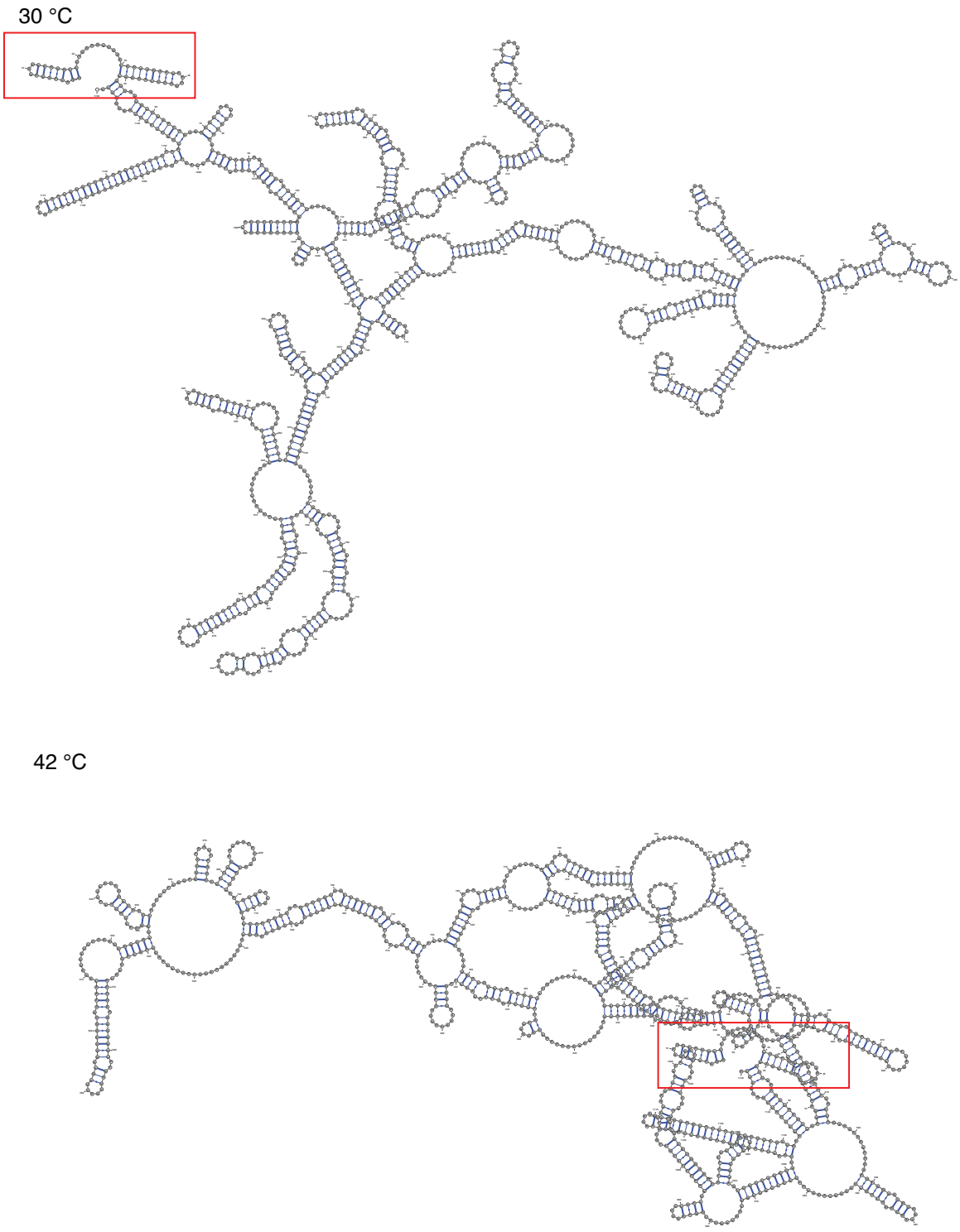


Figure D.9 Predicted *agsA* WT construct structures for full mRNA sequence *in vivo*. SHAPE constraints incorporated by RNAStructure covered 5' UTR (red box) and SFGFP leader sequence up until the site of reverse transcription primer binding.

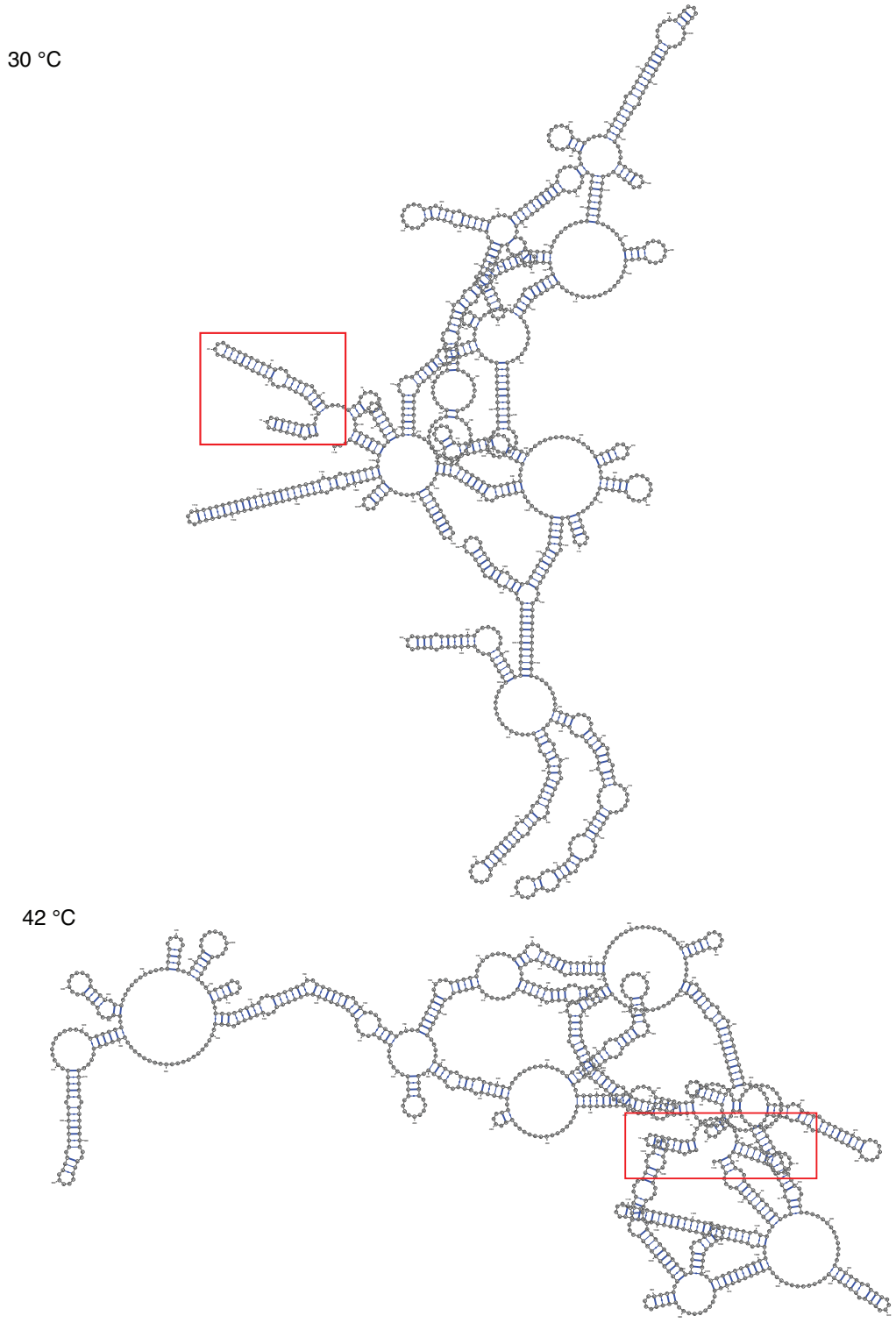


Figure D.10 Predicted *agsA* G21C construct structures for full mRNA sequence *in vivo*. SHAPE constraints incorporated by RNAStructure covered 5' UTR (red box) and SFGFP leader sequence up until the site of reverse transcription primer binding.

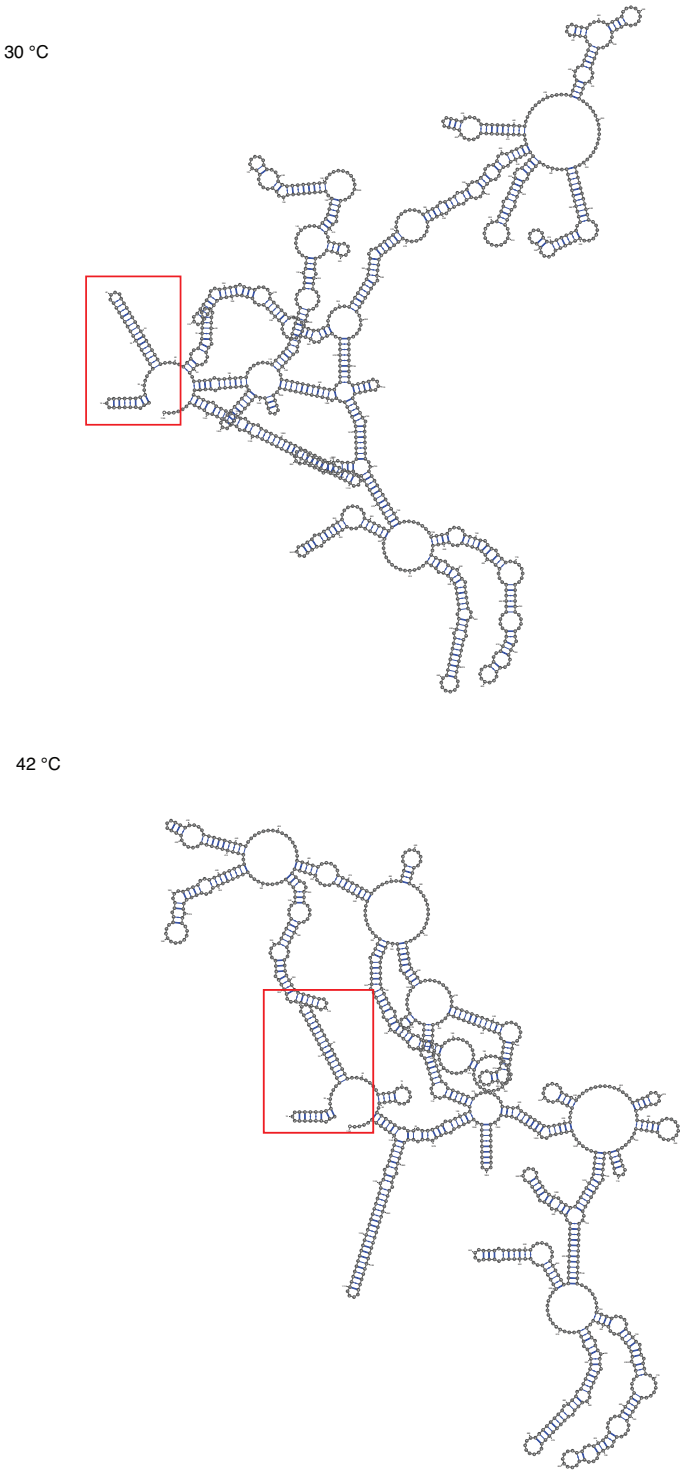


Figure D.11 Predicted *agsA* A29C construct structures for full mRNA sequence *in vivo*. SHAPE constraints incorporated by RNAStructure covered 5' UTR (red box) and SFGFP leader sequence up until the site of reverse transcription primer binding.

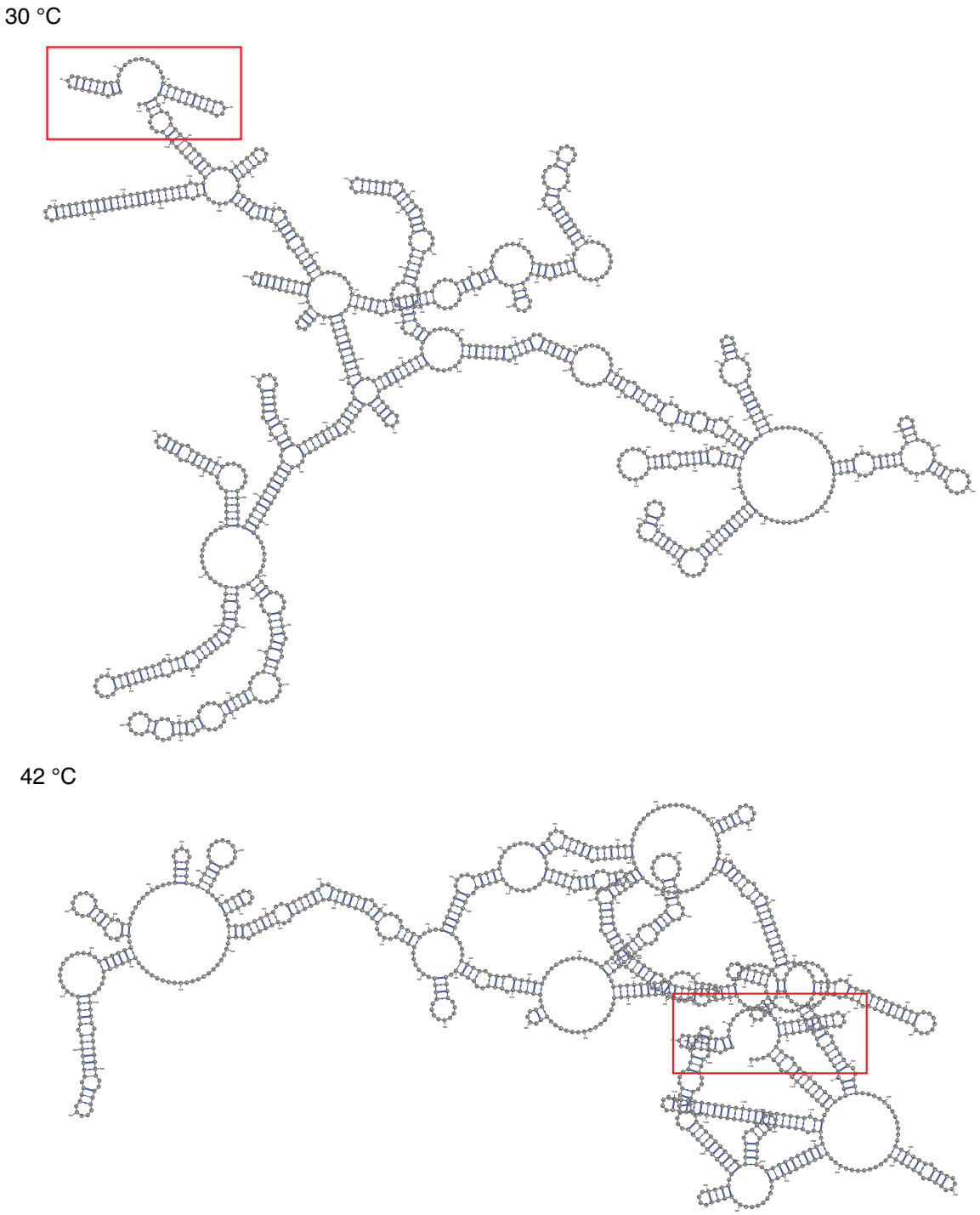


Figure D.12 Predicted *agsA* WT construct structures for full mRNA sequence *in vitro* in PURExpress without ribosomes. SHAPE constraints incorporated by RNAStructure covered 5' UTR (red box) and SFGFP leader sequence up until the site of reverse transcription primer binding.